

Contents lists available at [SciVerse ScienceDirect](http://SciVerse.Sciencedirect.com)

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Bio-PEPAd: A non-Markovian extension of Bio-PEPA

Giulio Caravagna^{a,*}, Jane Hillston^b^a Dipartimento di Informatica, Università di Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy^b Laboratory for Foundations of Computer Science, The University of Edinburgh, Edinburgh EH8 9AB, Scotland, United Kingdom

ARTICLE INFO

Article history:

Received 9 February 2011

Received in revised form 19 November 2011

Accepted 28 November 2011

Communicated by P. Degano

Keywords:

Bio-PEPA

Delay-as-duration

Non-Markovian process algebra

GSMPs

DSSA

DDEs

ABSTRACT

Delays in biological systems may be used to model events for which the underlying dynamics cannot be precisely observed, or to provide abstraction of some behavior of the system resulting in more compact models. In this paper, we enrich the stochastic process algebra Bio-PEPA, with the possibility of assigning delays to actions, yielding a new non-Markovian stochastic process algebra: Bio-PEPAd. This is a conservative extension meaning that the original syntax of Bio-PEPA is retained and the delay specification which can now be associated with actions may be added to existing Bio-PEPA models. The semantics of the firing of the actions with delays is the delay-as-duration approach, earlier presented in papers on the stochastic simulation of biological systems with delays. This semantics of the algebra is given in the Starting–Terminating style, meaning that the state and the completion of an action are observed as two separate events, as required by delays. We formally define the encoding of Bio-PEPAd systems in Generalized Semi-Markov Processes (GSMPs), as input for a Delay Stochastic Simulation Algorithm (DSSA) and as sets of Delay Differential Equations (DDEs), the deterministic framework for modeling of biological systems with delays. Finally, we prove theorems stating the relation between Bio-PEPA and Bio-PEPAd models. We end the paper with an example model of biological systems with delays to illustrate the approach.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The contribution of computer science to the interdisciplinary field of systems biology is to provide languages, tools and techniques for the description and analysis of complex biological systems. In particular, there exist many formal languages, either based on process algebras or term-rewriting systems, including the stochastic π -calculus [37,38,40], BioAmbients [39], the κ -calculus [20], the CLS [31,3,4], Bio-PEPA [15,14], BlenX [21] and LBS [35], to name but a few.

Biological systems can often be modeled at different abstraction levels. Specifically, a simple event in a model that describes the system at one level of detail may correspond to a rather complex network of events in a lower level description. The choice of the abstraction level of a model usually depends on the knowledge of the system and on the efficiency of the analysis tools to be applied to the model. Quantification of behavior is important for most biological models and increasingly it is being recognized that stochastic effects have a strong influence in many systems. As a result most of the process algebras or term-rewriting systems, have the ability to capture the dynamics of the system in a quantified way. Most commonly this is achieved by associating an exponentially distributed duration with reactions and giving the semantics of the language in terms of an underlying Continuous Time Markov Chain (CTMC) and/or a stochastic simulation algorithm based on Gillespie's algorithm [24]. In this paper, we are interested in systems where, in addition to this stochastically timed duration, there

* Corresponding author. Tel.: +39 050 2213143.

E-mail addresses: caravagn@di.unipi.it (G. Caravagna), jane.hillston@ed.ac.uk (J. Hillston).

is also a deterministically timed delay. In essence, this means that we make a distinction between the occurrence of the reaction, which happens at the end of the duration, and its effects, which become apparent after the delay.

Delays can appear in a biological system at any level of abstraction. We will illustrate this by informally discussing two very simple scenarios. Firstly, let us consider some complex dynamics (macro-event) decomposed in a series of sequential sub-events (micro-events): to explicitly model such dynamics we must have full quantitative information about all the sequential sub-events. This requirement implies that, if some information is missing then a full model cannot be described, and this is quite a common scenario in real modeling. If this is the case, we can think about a raw abstraction of this system by considering, instead of the micro-events, the single-step macro-event, assuming that we have enough information for it to be modeled. Delays come into play at this stage when the expected time for completion of the macro-event is known. Indeed, such information can be used to have a more precise model, as we shall see in the rest of the paper. Even though this is an abstraction of the exact model of the micro-events, this turns out to be the best that we can do in some situations. Moreover, there is another scenario in which delays turn out to be useful. Namely, when a system is too complex to analyze, then using a similar assumption to above, we can replace a collection of micro-events by a model with delay at the macro-event level. In this case, the delay is used as a model-reduction technique which makes the model smaller, and the analysis potentially feasible. Pragmatically, consider a series of n micro-events as

$$S_0 \xrightarrow{k_1} S_1 \xrightarrow{k_2} \dots \xrightarrow{k_n} S_n$$

transforming S_0 in S_n . In essence the first event is taken as the occurrence of the reaction as the change in S_0 can be observed, but the effect in terms of S_n is only apparent after a further delay. In reality this further delay will have a distribution which is dependent on the form of the reaction network between the occurrence and the event. This might be a Erlang distribution, a Coxian, or a more complex Phase Type distribution. However we would like to abstract this to a single step and we choose to represent this by a deterministic delay.

In mathematics, the modeling of biological systems with delays is mainly based on Delay Differential Equations (DDEs), a class of differential equations, obtained by generalizing Ordinary Differential Equations (ODEs), in which the derivative of the unknown function at a certain timepoint is given in terms of the values of the function at previous timepoints. This framework is very general and allows both simple (constant) and complex (variable or distributed) forms of delays to be modeled. Practically, DDEs have been used to describe biological systems in which events have a non-negligible duration [6,43] or in which a sequence of simple events is abstracted as a single complex event associated with a duration [42,17].

It is well-known that the analysis of ODEs can become imprecise due to the approximation introduced by representing discrete quantities by continuous variables when quantities are close to zero, and the same problem can arise in DDEs. Thus techniques for performing stochastic analysis of systems with delays have also been developed.

At a theoretical level, the introduction of constant delays means that the semantics of actions result in a non-Markovian stochastic process. We will show that in general the resulting process is a Generalized Semi-Markov Process (GSMP) [26,19,16,10,11]. Such processes are discrete processes, where the embedded state process is a Markov chain, but the time between jumps is a random variable of arbitrary distribution, which may be dependent on the two states between which the move will be made. If, in each state, there is a single jump event, then the process is a Semi-Markov Process, in contrast to a GSMP which may have more than one jump event concurrently running in each state. If in a Semi-Markov Process the time between jumps is exponentially distributed (i.e. memoryless) then it is a Continuous-Time Markov Chain (CTMC). GSMPs have been used to give a stochastic process description of a large class of discrete-event simulations [26,16].

From a pragmatic point of view, techniques for simulation of biological systems with delays have been defined. The Delay Stochastic Simulation Algorithms (DSSAs) [5,8,1,2,12], often exploiting Gillespie's Stochastic Simulation Algorithm (SSA) of chemical reactions [24], permit the computation of a time-trace of the non-Markovian stochastic process underlying a model with delays. These algorithms permit different interpretations of delays and, for some of these, it has been shown that GSMPs underlie the simulated system [12]. Moreover, the relationship between DSSAs and DDEs has been outlined by means of *Delay Chemical Master Equations* [5,12], the extension with delays of the Master Equations logically connecting the SSA and ODEs [24]. Among these interpretations, it is possible to have a *delay-as-duration* approach to the firing of reactions, or a *purely delayed* one. In the former [5,1,2], the reactants are removed at the beginning of a reaction and the products are added at its end, namely after the delay plus an exponentially distributed time quantity. In this sense, during the time of firing of the reaction, the reactants will not be able to take part in other reactions. According to [1,2], for some biological systems it is necessary that reactants involved in a reaction with delay can have other interactions while waiting for the delay to complete. In the latter interpretation, the purely delayed approach, the reactants involved in a reaction can have other interactions during the firing of the reaction itself. This interpretation is more complex to adopt, both at the level of DSSAs and formal languages, as discussed in [12].

In this paper, we define a process algebra for the modeling of biological systems with delays. More precisely, we use constant delays in the DDEs and, for the DSSAs, we take the delay-as-duration approach presented in [1,12]. These restrictions are reasonable since they permit us to have a simple algebra obtained by extending a well-known one, Bio-PEPA [14,15]. Moreover, our work provides scope for later versions in which this algebra may be extended to more complex forms and interpretations of delays.

Process algebras dealing with general distributions (e.g. Erlang, Coxian, Hyperexponential or Gamma) have been defined to model either generic concurrent systems [18,29,9,11,30] or biological systems [33,34]. Our work differs from such process

algebras since we focus on a specific type of general distribution, which is inspired by the mathematical modeling of biological systems with delays. To this extent, we want to define a process algebra where the semantics of actions is driven by a well-known semantics of the firing of a chemical reaction with delays, i.e. the delay-as-duration approach. Secondly, we want this algebra to provide operators which are tailored to model biological systems. In fact, in this work we extend Bio-PEPA, a stochastic process algebra for the modeling and the analysis of biochemical networks. Bio-PEPA is based on PEPA [27], a process algebra originally defined for the performance analysis of computer systems, and extends it in order to handle typical features of biochemical networks, such as stoichiometry and various types of kinetic laws. A main feature of Bio-PEPA is the ability to support various types of analysis. In particular, Bio-PEPA models can be analyzed using stochastic simulation based on Gillespie's SSA [24] and steady state analysis can be performed on the Continuous-Time Markov Chain underlying the semantics of a model. Furthermore, Bio-PEPA models can be translated into equivalent deterministic models based on ODEs and, finally, they can be model checked using the PRISM [28,41] model checker. The Bio-PEPA modeling paradigm is *processes-as-species* rather than *processes-as-molecules*, as in the Stochastic π -calculus [38]. This choice may permit a model with a smaller state space, making analysis more tractable.

In this paper we enrich the stochastic process algebra Bio-PEPA with the possibility of assigning delays to actions, yielding the definition of a new non-Markovian process algebra: Bio-PEPA_d. The new algebra is based on the same syntax as Bio-PEPA, hence the definition of Bio-PEPA_d systems with delays can be easily obtained by adding, to a Bio-PEPA system of the target model, the delay specifications. A key feature of Bio-PEPA, namely the notion of concentration level for species used to tackle state space explosion, is also present in Bio-PEPA_d. The semantics of the new algebra is given in the Starting–Terminating (ST) style [25,10], which allows us to observe the start and the completion of an action as two separate events, as required by delays. By using the ST style to give the semantics of Bio-PEPA_d we also outline a clear relation between the algebra and the underlying GSMPs. Moreover, there is a strong relationship with the biology since the start corresponds to the occurrence of a reaction while the termination corresponds to its effects being apparent. The clear operational relationship between Bio-PEPA and Bio-PEPA_d gives us the opportunity to prove theorems on the correspondence between the semantics of the two languages and the relation between the mathematical structures underlying the models.

Following previous work on Bio-PEPA analysis, we outline how to automatically translate a system in a set of DDEs, how to perform stochastic simulation of Bio-PEPA_d systems using a DSSA and how to encode Bio-PEPA_d systems in GSMPs. During the presentation of the definitions we use examples to clarify the approach and, at the end of the paper we encode in Bio-PEPA_d a well-known model of the cell cycle with delays where the passage of cells from different phases of the cell cycle is modeled by a delay. This model is then translated into a set of DDEs and in a set of reactions; the former matches the original deterministic definition of the model [42], and the latter its original stochastic definition [1].

The paper is structured as follows: in Section 2 we recall the definitions of Bio-PEPA that we maintain in the definition of Bio-PEPA_d. In Section 3 we separately introduce the syntax and the semantics of the new language. We discuss the automatic translation of Bio-PEPA_d models into DDEs, input for the DDA and GSMPs in Section 4. In Section 5 we prove results stating the operational connection between Bio-PEPA and Bio-PEPA_d. Finally, in Section 6 a well-known model of the cell cycle with a delay is presented in Bio-PEPA_d and in Section 7 conclusions and future work are discussed. This paper significantly extends our original work on Bio-PEPA_d which appeared in [13].

2. Bio-PEPA

Bio-PEPA [14,15] is a stochastic process algebra, based on PEPA [27], for the modeling and the analysis of biochemical networks. The operators of this algebra are designed to make the description of biochemical networks easy. Indeed, features such as stoichiometry of reactions and general kinetic laws can be conveniently captured in Bio-PEPA models. Furthermore, as already mentioned in the previous section, the algebra supports multiple analysis techniques for the defined models. Stochastic simulations, steady state analysis of the CTMC, automatic translation into sets of deterministic ODEs and, finally, model checking analysis, can be performed on Bio-PEPA models.

The processes-as-species modeling paradigm of Bio-PEPA supports a compact state space representation and, consequently, a model whose analysis is more likely to be feasible. A model is described by sequential components representing species, and by a model component representing their possible interactions.

In this section we recall the parts of the definition of Bio-PEPA that we will use to define Bio-PEPA with delays. We assume a set of action types \mathcal{A} and we start by recalling the syntax of the processes.

Definition. Bio-PEPA *species* and *processes* are defined by the following grammar:

$$\begin{aligned} S &::= (\alpha, \kappa)op\ S \mid S + S \mid C \\ P &::= P \boxtimes_{\mathcal{L}} P \mid S(l) \end{aligned}$$

where $op \in \{\downarrow, \uparrow, \odot, \oplus, \ominus\}$, $\alpha \in \mathcal{A}$, \mathcal{L} is a set of actions and $l, \kappa \in \mathbb{N}$. We denote with \mathcal{S} the set of all possible species specifications, and we denote with \mathcal{P} the set of all possible well-formed Bio-PEPA processes, as defined in [14].

The components S and P represent species and their possible interactions, respectively. The element C is used to define constant processes.

$$\begin{array}{c}
(\alpha, \kappa) \downarrow S(l) \xrightarrow{(\alpha, [S: \downarrow(l, \kappa)])}_c (\alpha, \kappa) \downarrow S(l - \kappa) \quad \kappa \leq l \leq N \\
(\alpha, \kappa) \uparrow S(l) \xrightarrow{(\alpha, [S: \uparrow(l, \kappa)])}_c (\alpha, \kappa) \uparrow S(l + \kappa) \quad 0 \leq l \leq N - \kappa \\
(\alpha, \kappa) \oplus S(l) \xrightarrow{(\alpha, [S: \oplus(l, \kappa)])}_c (\alpha, \kappa) \oplus S(l) \quad 0 < l \leq N \\
(\alpha, \kappa) op S(l) \xrightarrow{(\alpha, [S: op(l, \kappa)])}_c (\alpha, \kappa) op S(l) \quad 0 \leq l \leq N, op \in \{\odot, \ominus\} \\
\frac{S_1(l) \xrightarrow{(\alpha, w)}_c S'_1(l',)}{(S_1 + S_2)(l) \xrightarrow{(\alpha, w)}_c S'_1(l')} \quad \frac{S_2(l) \xrightarrow{(\alpha, w)}_c S'_2(l')}{(S_1 + S_2)(l) \xrightarrow{(\alpha, w)}_c S'_2(l')} \\
\frac{S(l) \xrightarrow{(\alpha, w)}_c S'(l') \quad C \stackrel{def}{=} S}{C(l) \xrightarrow{(\alpha, w)}_c S'(l')} \quad \frac{P_1 \xrightarrow{(\alpha, w)}_c P'_1 \quad \alpha \notin \mathcal{L}}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha, w)}_c P'_1 \boxtimes_{\mathcal{L}} P_2} \\
\frac{P_1 \xrightarrow{(\alpha, w_1)}_c P'_1 \quad P_2 \xrightarrow{(\alpha, w_2)}_c P'_2 \quad \alpha \in \mathcal{L}}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha, w_1 @ w_2)}_c P'_1 \boxtimes_{\mathcal{L}} P'_2} \quad \frac{P_2 \xrightarrow{(\alpha, w)}_c P'_2 \quad \alpha \notin \mathcal{L}}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha, w)}_c P_1 \boxtimes_{\mathcal{L}} P'_2} \\
\frac{P \xrightarrow{(\alpha, w)}_c P' \quad r_\alpha = f_\alpha[w, \mathcal{N}, \mathcal{K}]h^{-1}}{\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P \rangle \xrightarrow{(\alpha, r_\alpha)}_s \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P' \rangle}
\end{array}$$

Fig. 1. The Bio-PEPA capability (\rightarrow_c) and stochastic (\rightarrow_s) relations.

Bio-PEPA actions are used to model the events (i.e. the reactions) happening in the biological systems we model. The prefix terms in this algebra contain information about the role of the species in the actions. In particular, for $(\alpha, \kappa) op S$ we have that (α, κ) is the prefix, where $\alpha \in \mathcal{A}$ is the action type and κ is the stoichiometry coefficient of the species in the reaction. The prefix combinator “ op ” indicates the role of the species in the reaction. In particular, \downarrow indicates a reactant, \uparrow a product, \oplus an activator, \ominus an inhibitor and \odot a generic modifier. The species can appear in a sum $S_1 + S_2$, whose meaning is the classical “choice” of process algebras.

Following the processes-as-species paradigm, in Bio-PEPA a discrete concentration level l is associated with each species. During the simulation of a system, the concentration of a species S , denoted by $S(l)$ ranges over $\{0, \dots, N_S\}$, where N_S is its maximum level of concentration statically defined to bound the population size. Also, a fixed step size h for all the species is defined. This means that changing the concentration level of a species by one implies a change in h units of concentration of that species, considering the actual counts and volumes. The granularity, as well as the rate functions, are defined in terms of the step size h of the concentration intervals. This choice permits us to deal with incomplete information in the exact number of elements, and leads to a reduction of the state space as there are less states for each component.

Bio-PEPA supports multiway synchronization, i.e. synchronization can involve more than two components. This makes it easy to model n -ary reactions, whose modeling in dyadic process algebras is not trivial. The term $P_1 \boxtimes_{\mathcal{L}} P_2$ denotes cooperation between P_1 and P_2 over the *cooperation set* \mathcal{L} , which determines those activities on which the cooperands are forced to synchronize. For action types not in \mathcal{L} , the components proceed independently and concurrently with their enabled activities.

A Bio-PEPA model specification is given in terms of a system, where a system is defined as follows.

Definition. A Bio-PEPA system \mathcal{P} is a 6-tuple $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P \rangle$ where:

- \mathcal{V} is the set of *compartments*;
- \mathcal{N} is the set of *quantities* describing each species;
- \mathcal{K} is the set of *parameter* definitions;
- \mathcal{F} is the set of *functional rate* definitions;
- $Comp$ is the set of *sequential component* definitions;
- P is the *initial process* definition.

We denote the set of all possible Bio-PEPA systems as \mathcal{R} . Notice that in Bio-PEPA the kinetic characteristics of the actions are not specified in the syntax of processes as in other calculi but, instead, they are separately represented in the notation of system. Indeed, in this definition the information about rates is given by \mathcal{F} and that about kinetic constants is given by \mathcal{K} , while the initial process definition is P .

The semantics of Bio-PEPA is given by a Structural Operational Semantics (SOS) [36], similar to the one for PEPA, and is given in Fig. 1. The semantics is based on a capability relation $\rightarrow_c \subseteq \mathcal{P} \times \Theta \times \mathcal{P}$ where

$$\Theta = \{(\alpha, w) \mid \alpha \in \mathcal{A}, w \in W\}.$$

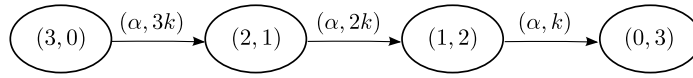


Fig. 2. The SLTS for the Bio-PEPA example.

With W we denote the set of lists defined by the grammar

$$w ::= [S : op(l, \kappa)] \mid w@w$$

where $S \in \mathcal{S}$, $op \in \{\downarrow, \uparrow, \odot, \oplus, \ominus\}$, $l, \kappa \in \mathbb{N}$, and $@$ the classical concatenation operator on lists [32]. Labels from Θ contain the information needed in order to evaluate the functional rate; the capability relation supports the derivation of quantitative information and is auxiliary to a stochastic relation $\rightarrow_s \subseteq \mathcal{R} \times \Gamma \times \mathcal{R}$ where

$$\Gamma = \{(\alpha, r_\alpha) \mid \alpha \in \mathcal{A}, r_\alpha \in \mathbb{R}^+\}.$$

The stochastic relation associates the rate r_α with the action α performed. The rates are obtained by evaluating the functional rate associated with the action, divided by the step size, and by using the quantitative information derived from the capability relation, as explained in [14]. The use of two relations allows for the association of the rate with the last step of the derivation representing a given reaction, going from a capability transition to a stochastic transition. This makes it easier to derive the rate in the appropriate way, especially in the case of general kinetic laws different from mass-action. By using the stochastic relation it is possible to define the semantics of Bio-PEPA as a *Stochastic Labeled Transition System* (SLTS). For the precise definitions and explanations of the components of a Bio-PEPA system, as well as for further comments on the SOS of Bio-PEPA, we refer to [14].

A Bio-PEPA toy example

In order to clarify modeling with Bio-PEPA we present a toy example. We model a transformation event from one element of species A to one element of species B . Transformation happens at a rate k and obeys a mass-action kinetic law. Such a model is constituted by a single reaction channel of the form $A \xrightarrow{k} B$. The initial state contains three elements of species A and no elements of species B ; formally it is described by the 2-dimensional vector $\mathbf{x}_0 = (3, 0)^T$.

The Bio-PEPA processes modeling the species are

$$A \stackrel{\text{def}}{=} (\alpha, 1)\downarrow A \quad B \stackrel{\text{def}}{=} (\alpha, 1)\uparrow B$$

where α is the action corresponding to the reaction, the functional rates are defined according to the mass-action kinetics, namely by defining $f_\alpha = f_{MA}(k')$ when $k' = k$. The Bio-PEPA process $A \underset{\{\alpha\}}{\bowtie} B$ describes the interaction of the components and represents the fact that the two processes synchronize to perform action α , the transformation.

By considering levels we assume the species to have some maximum levels N_A and N_B , which in this case are $N_A = N_B = 3$. The initial levels of concentrations are described by the vector \mathbf{x}_0 , and the initial configuration of the process is the following

$$A(3) \underset{\{\alpha\}}{\bowtie} B(0).$$

The components of the system in which this process is embedded are

$$\begin{aligned} \mathcal{V} &= \{\text{cell} : 1\} & \mathcal{K} &= \{k' = k\} \\ \mathcal{N} &= \{A \text{ in cell} : N_A = 3, h_A = 1; \\ &\quad B \text{ in cell} : N_B = 3, h_B = 1\} & \mathcal{F} &= \{f_\alpha = f_{MA}(k')\} \\ \text{Comp} &= \{A \stackrel{\text{def}}{=} (\alpha, 1)\downarrow A, B \stackrel{\text{def}}{=} (\alpha, 1)\uparrow B\} & P &= A(3) \underset{\{\alpha\}}{\bowtie} B(0). \end{aligned}$$

By applying the Bio-PEPA stochastic relation to the system with this initial process we obtain the unique possible evolution of the system described by a finite SLTS, as expected. Also, there is a one-to-one correspondence between the states and the transitions of the SLTS, on the one hand, and of the CTMC underlying the stochastic process described by the model, on the other hand.

A graphical representation of the state-transitions for the process is given in Fig. 2. In that figure, all the states are represented as circles where the notation (n_1, n_2) represents the discrete levels of concentration n_1 and n_2 of the species A and B , respectively. All the arrows represent stochastic derivations of the whole system, where the labels are exactly those computed by that relation. So, for instance, by using the label $[A : (3, 1)\downarrow][B : (3, 0)\uparrow]$ exhibited by the capability relation the rate $3k$ is evaluated. For a detailed description of how the lists appearing as labels are used to derive the rates we refer to [14].

As expected, this system, starting from the initial configuration $\mathbf{X}(t_0) = \mathbf{x}_0$ at some initial time t_0 , eventually reaches the final state $(0, 3)$, which corresponds to the process $A(0) \underset{\{\alpha\}}{\bowtie} B(3)$ and to the vector $(0, 3)^T$.

3. Bio-PEPAd: Bio-PEPA with delays

In the following subsections we separately present the syntax and the semantics of Bio-PEPA with delays (Bio-PEPAd).

3.1. Syntax and process configurations

Processes of Bio-PEPAd are defined by the same syntax as Bio-PEPA processes, hence it will be possible to easily extend a Bio-PEPA system into one with delays.

In Bio-PEPA the general kinetic information is specified separately from the syntax of processes. The delays, which are also properties of the actions which can be performed, are similarly represented separately in Bio-PEPAd. Indeed, they are defined by functions belonging to the family

$$\left\{ \sigma : \mathcal{A} \rightarrow \mathbb{R}^+ \right\} \in \Delta \quad (1)$$

such that $\sigma(\alpha)$ denotes the delay of action $\alpha \in \mathcal{A}$. From the biological perspective, the choice of using σ as a function on actions to specify the delays implies that every participant in an action α will have the same delay $\sigma(\alpha)$, which is sound since for each species involved in the reaction modeled by α the delay is unique. For the sake of simplicity we assume all the actions to have a non-zero delay, the combination of delayed and non-delayed actions can be defined in a natural way by merging the results we present here together with those in [14]. A Bio-PEPAd system is defined as an extension of a Bio-PEPA one as follows.

Definition. A Bio-PEPAd system is a 7-tuple $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, P \rangle$ where:

- $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp} \rangle$ is a Bio-PEPA system;
- $\sigma \in \Delta$ is a function used to specify the delays of the actions.

We denote with $\tilde{\mathcal{P}}$ the set of all possible Bio-PEPAd systems.

In order to define the semantics of Bio-PEPAd we define a notion of process configuration.

Definition. Bio-PEPAd *process configurations* are defined by the following syntax:

$$\begin{aligned} C_S &::= (\alpha, \kappa) op C_S \mid C_S + C_S \mid C \\ C_P &::= C_P \boxtimes_{\mathcal{L}} C_P \mid C_S(l, L) \end{aligned}$$

where L is a list of 4-tuples $(l', \kappa', \alpha', op')$ with $l, \kappa \in \mathbb{N}$, $\alpha \in \mathcal{A}$ and $op \in \{\downarrow, \uparrow, \odot, \oplus, \ominus\}$. We denote with \mathcal{C} the set of all well-formed process configurations.

The notion of a well-formed process configuration is straightforward; any process configuration is well-formed if, by removing the list L , its corresponding Bio-PEPA process is well-formed. For clarity, in the following we denote a generic process configuration as $S(l, L)$.

In contrast to processes, which describe possible actions, process configurations also describe which actions are running. A species $S(l, L)$ is a species with a discrete level of concentration l , like the species $S(l)$ in Bio-PEPA, but which is currently involved in the actions with delay described by the list L . In particular, if the list L contains an entry (l', κ, α, op) , this means that there are κ levels of concentration of species S involved in a currently running action α which fired when the discrete level of concentration of species S was l' , its role in this instance of the action is described by op . For instance, a species $S(3, [(2, 1, \alpha, \uparrow)])$ is a species with current concentration level 3, involved in a scheduled action α , started when its concentration level was 2, which is going to increase by 1 its concentration level when completed.

Consequently, L is to be considered as a view of the scheduling list used in the algorithms described in [1] for simulating stochastic models with delays. More precisely, L is a view of only the scheduled events which involve elements of species S .

In order to define the semantics of Bio-PEPAd, it is necessary to define some auxiliary functions for manipulating the scheduling lists L . We denote by \mathcal{R} the domain of all the possible tuples of the form (l, κ, α, op) , and with \mathcal{R}^* all the possible lists built over \mathcal{R} by using the monoidal product $::$, hence $L \in \mathcal{R}^*$. We start by defining four functions

$$\begin{aligned} \text{pick} : \mathcal{A} \rightarrow \mathcal{R}^* \rightarrow \mathcal{R} & & \text{del} : \mathcal{A} \rightarrow \mathcal{R}^* \rightarrow \mathcal{R}^* \\ \text{prod} : \mathcal{R}^* \rightarrow \mathcal{R}^* & & \text{pend} : \mathcal{R}^* \rightarrow \mathbb{N} \end{aligned}$$

whose formal definition is given in Table 1.

Function *pick* extracts the first scheduled event with a given action name from the list L . The function value *pick* α L is \perp if no entries of action α exist in L (i.e. no actions α are currently running); otherwise, it is the first occurrence obtained by a left-to-right recursive scan of L , if any. Notice that we assume the syntactic priority of pattern matching.

Function *del* is used to modify a list such that *del* α L is a new list obtained by removing the first, if any, occurrence of an action α obtained by a left-to-right scan of L . As this is an event list, the ordering of insertion of the tuples determines their ordering for extraction. The functions *pick* and *del*, together with the classical append function on lists, namely function *@*,

Table 1Formal functional-style [32] definitions of auxiliary functions *pick*, *del*, *prod* and *pend*.

$pick\ \alpha\ L = \text{match } L \text{ with}$	$del\ \alpha\ L = \text{match } L \text{ with}$
$ [] \rightarrow \perp;$	$ [] \rightarrow [];$
$ (l, \kappa, \alpha, op) :: xs \rightarrow (l, \kappa, \alpha, op);$	$ (l, \kappa, \alpha, op) :: xs \rightarrow xs;$
$ x :: xs \rightarrow pick\ \alpha\ xs.$	$ x :: xs \rightarrow x :: del\ \alpha\ xs.$
$prod\ L = \text{match } L \text{ with}$	$pend\ L = \text{match } L \text{ with}$
$ [] \rightarrow [];$	$ [] \rightarrow 0;$
$ (l, \kappa, \alpha, \uparrow) :: xs \rightarrow (l, \kappa, \alpha, \uparrow) :: prod\ xs;$	$ (l, \kappa, \alpha, op) :: xs \rightarrow \kappa + pend\ xs.$
$ x :: xs \rightarrow prod\ xs.$	

will be used to implement a First-In First-Out (FIFO) policy for insertion and extraction of elements in L . As an example, given a list $l_1 \equiv (2, 1, \alpha, \downarrow) :: l_2$ the functions are such that $pick\ \alpha\ l_1 = (2, 1, \alpha, \downarrow)$ and $del\ \alpha\ l_1 = l_2$, namely the function *del* removes the entry computed by the function *pick*, when applied with the same parameter.

As in Bio-PEPA we want to keep the state representation of the models finite by using some constraints for the starting of actions. Thus, let us denote the scheduled actions in which the species S is involved as a product by $prod\ L$. The species $S(l, L)$ is currently involved in the delayed actions as follows: for the scheduled actions in $prod\ L$ it is involved as a product, and for the other ones it is involved either as a reactant, a modifier, an activator or an inhibitor.

Finally, let us denote by *pend* the function computing how many levels of concentration are involved in all the actions described in its input list, regardless of the role of the species in the scheduled event. By following the delay-as-duration of approach [1] in the interpretation of the delays this implies that, for species S , there are exactly $pend\ prod\ L$ levels of concentration of species S which are currently waiting for their delay to expire before becoming available in the species S . These two functions will be used to define the constraints to keep the state space finite, as presented in the next subsections. For instance, given the list l_1 previously defined, the functions are such that $prod\ l_1 = prod\ l_2$ and $pend\ prod\ l_1 = pend\ l_2$ since the first entry of l_1 is discarded.

A Bio-PEPAD system specification is typically given in terms of a process $P \in \mathcal{P}$ whose semantics is given in terms of its equivalent process configuration $P_C \in \mathcal{C}$. Intuitively, we want the initial term P to be modified in the corresponding initial configuration P_C where every species declaration $S(l_0, [])$ in P_C is such that $S(l_0)$ is in P . The initial process configuration is obtained by adding an empty scheduling list to each species because, in the initial configuration, there are no instances of actions with delay currently running. Formally, we define, by structural recursion on the process syntax a function $\mu : \mathcal{P} \rightarrow \mathcal{C}$ such that

$$\begin{aligned} \mu((\alpha, \kappa)op\ S) &= (\alpha, \kappa)op\ S & \mu(P_1 \boxtimes_{\mathcal{L}} P_2) &= \mu(P_1) \boxtimes_{\mathcal{L}} \mu(P_2) \\ \mu(S_1 + S_2) &= S_1 + S_2 & \mu(S(l)) &= S(l, []). \end{aligned}$$

As expected the function is such that the process $S(l_1) \boxtimes_{\mathcal{L}_1} S(l_2) \boxtimes_{\mathcal{L}_2} S(l_3)$ is transformed into the configuration $S(l_1, []) \boxtimes_{\mathcal{L}_1} S(l_2, []) \boxtimes_{\mathcal{L}_2} S(l_3, [])$. Notice that empty scheduling lists are associated with Bio-PEPAD processes. This represents a specific initial state for the stochastic process underlying Bio-PEPAD, as we will discuss in Section 4.3.

We augment the definition of Bio-PEPAD systems to 7-tuples of the form $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, \sigma, P_C \rangle$ where P_C is a process configuration of a process. In the following, we may use the notation P to refer to either a process or a process configuration; it will be clear from the context to which of them we are referring. We denote the extended set of all Bio-PEPAD systems with process configurations as \mathcal{P} .

Similarly to Bio-PEPA where the SOS is defined by means of two relations, in this algebra the SOS, given in a Starting-Terminating (ST) style, is defined by means of three relations that we present in the following section.

3.2. A structural operational semantics for Bio-PEPAD

In the following subsections we define a *start relation* on process configurations which, in the same style as the Bio-PEPA capability relation, contains the quantitative information needed to evaluate the functional rates and modifies the process configurations to model the start of an action. Also, we define a *completion relation* on process configurations which describes the termination of an action. Finally, along the lines of the stochastic relation in Bio-PEPA, we define a *stochastic relation* for Bio-PEPAD systems, based on the start and completion relations, which associates rates with transitions. We remark that the separate relations are used to give a semantics of the start (occurrence) and completion (effect) of an action in accordance with the delay-as-duration approach [12]. As noted earlier, we assume only systems where all the actions are delayed. In contrast, we could have adopted a semantics which allows delayed and non-delayed actions to be mixed, such as the racing timed transition system of Markovski [30]. This remains a possible topic for future work.

$$\begin{array}{l}
(\alpha, \kappa) \downarrow S(l, L) \xrightarrow{(\alpha^+, [S: \downarrow(l, \kappa)])}_{st} S(l - \kappa, L @ [(l, \kappa, \alpha, \downarrow)]) \quad \kappa \leq l \leq N \\
(\alpha, \kappa) \uparrow S(l, L) \xrightarrow{(\alpha^+, [S: \uparrow(l, \kappa)])}_{st} S(l, L @ [(l, \kappa, \alpha, \uparrow)]) \quad 0 \leq l + \text{pend prod } L \leq N \\
(\alpha, \kappa) \oplus S(l, L) \xrightarrow{(\alpha^+, [S: \oplus(l, \kappa)])}_{st} S(l, L @ [(l, \kappa, \alpha, \oplus)]) \quad \kappa \leq l \leq N \\
(\alpha, \kappa) op S(l, L) \xrightarrow{(\alpha^+, [S: op(l, \kappa)])}_{st} S(l, L @ [(l, \kappa, \alpha, op)]) \quad 1 \leq l \leq N, \quad op \in \{\odot, \ominus\} \\
\hline
\frac{S_1(l, L) \xrightarrow{(\alpha^+, w)}_{st} S'_1(l', L')}{(S_1 + S_2)(l, L) \xrightarrow{(\alpha^+, w)}_{st} S'_1(l', L')} \quad \frac{S_2(l, L) \xrightarrow{(\alpha^+, w)}_{st} S'_2(l', L')}{(S_1 + S_2)(l, L) \xrightarrow{(\alpha^+, w)}_{st} S'_2(l', L')} \\
\frac{S(l, L) \xrightarrow{(\alpha^+, w)}_{st} S'(l', L') \quad C \stackrel{def}{=} S(l, L)}{C \xrightarrow{(\alpha^+, w)}_{st} S'(l', L')} \quad \frac{P_1 \xrightarrow{(\alpha^+, w)}_{st} P'_1 \quad \alpha \notin \mathcal{L}}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha^+, w)}_{st} P'_1 \boxtimes_{\mathcal{L}} P_2} \\
\frac{P_1 \xrightarrow{(\alpha^+, w_1)}_{st} P'_1 \quad P_2 \xrightarrow{(\alpha^+, w_2)}_{st} P'_2 \quad \alpha \in \mathcal{L}}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha^+, w_1 @ w_2)}_{st} P'_1 \boxtimes_{\mathcal{L}} P'_2} \quad \frac{P_2 \xrightarrow{(\alpha^+, w)}_{st} P'_2 \quad \alpha \notin \mathcal{L}}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha^+, w)}_{st} P_1 \boxtimes_{\mathcal{L}} P'_2}
\end{array}$$

Fig. 3. The start relation $\rightarrow_{st} \subseteq \mathcal{C} \times \Theta^+ \times \mathcal{C}$.

The start relation

This relation contains the quantitative information to compute rates of starting actions. Also, this relation modifies the process configuration to model the starting of an action.

The start relation is $\rightarrow_{st} \subseteq \mathcal{C} \times \Theta^+ \times \mathcal{C}$ where Θ^+ is an adaptation of Θ with labels of the form (α^+, w) ; the list w is of the same type as the one exhibited as a label by the capability relation of Bio-PEPA. The Bio-PEPAD start relation is defined as the minimal relation satisfying the rules presented in Fig. 3.

Formally, if a species $(\alpha, \kappa) \downarrow S(l, L)$ is involved as reactant in an action, then by following the delay-as-duration approach [1] its concentration level is decreased by κ . Differently, in the case of a species involved as a product, its concentration level is not changed because, as previously stated, this relation models the starting and not the completion of an action with delay. In the case of a species taking part in the reaction as a modifier, an inhibitor or an activator, its concentration level is unchanged, as expected. Independently of the role of a species, its scheduling list L is modified to record that some of its levels of concentration are currently performing action α . Notice that, in order to maintain the FIFO property on the scheduling list L , we simply use the append function $@$. This is possible because of the processes-as-species paradigm and the use of fixed deterministic delays. More precisely, it is required by the delay-as-duration approach that two instances of the same action complete according to their starting order. It is easy to see that this is certainly true in a framework where delays are deterministic. Differently, if they had been stochastic specific assumptions on the dependences of the delays would have been necessary. In the most general case unrelated stochastic delays would lead to action completing in any possible order, thus requiring more complex strategies to handle scheduling lists. In our case, the use of scheduling lists local to species guarantees that, as required, the two instances will appear sequentially and, hence, will complete sequentially. Practically, the order of appearance in the lists makes instances of the same action distinguishable.

We use constraints on the levels to have a finite state space as in Bio-PEPA. The constraints for starting the actions are the same as those in Bio-PEPA except the one for the products. In particular, the constraints which must be satisfied by a species S with level l and scheduling list L to fire an action as a product is, as expected, $0 \leq l + \text{pend prod } L \leq N$, if N is its maximum level. Intuitively, this means that the levels of concentration in the state, l , plus those which are already scheduled to be produced, $\text{pend prod } L$, must not exceed the capacity threshold N .

The starting of the action α , in the style of the ST semantics, is denoted by the action symbol α^+ , exhibited as a label for all the start derivations. The composition of the derivations of this relation is straightforward.

Some further considerations and comparisons with Bio-PEPA are useful. Firstly, when the actions have no delay as in Bio-PEPA, whenever an action fires, the changes in the process are immediately visible in a one-step derivation, since the Bio-PEPA capability relation modifies the process according to the action. In this algebra, as the instants at which an action starts and terminates are detached, then the start relation modifies the process to represent only the starting of the action. Indeed, another relation, which does not exist in the semantics of Bio-PEPA, will model the termination of a currently running action.

Secondly, by comparing the algorithm presented in [1] and the definition of this relation, it is clear that the modification of the process to reflect the starting of an action corresponds to scheduling of the reaction in the scheduling list, once the reactants have been consumed.

The completion relation

This relation is used to model the completion of an action with delay which is currently running. Again, this relation involves quantitative information needed to re-compute the functional rate of the action at the moment in which it started.

$$\begin{array}{c}
\frac{\text{pick } \alpha L = (l, \kappa, \alpha, \uparrow)}{S(l', L) \xrightarrow{(\alpha^-, [S: \uparrow(l, \kappa)])}_{co} S(l' + k, \text{del } \alpha L)} \quad \frac{\text{pick } \alpha L = (l, \kappa, \alpha, \text{op}) \text{ op} \in \{\downarrow, \odot, \oplus, \ominus\}}{S(l', L) \xrightarrow{(\alpha^-, [S: \text{op}(l, \kappa)])}_{co} S(l', \text{del } \alpha L)} \\
\\
\frac{S_1(l, L) \xrightarrow{(\alpha^-, w)}_{co} S'_1(l', L')}{(S_1 + S_2)(l, L) \xrightarrow{(\alpha^-, w)}_{co} S'_1(l', L')} \quad \frac{S_2(l, L) \xrightarrow{(\alpha^-, w)}_{co} S'_2(l', L')}{(S_1 + S_2)(l, L) \xrightarrow{(\alpha^-, w)}_{co} S'_2(l', L')} \\
\\
\frac{S(l, L) \xrightarrow{(\alpha^-, w)}_{co} S'(l', L') \quad C \stackrel{\text{def}}{=} S(l, L)}{C \xrightarrow{(\alpha^-, w)}_{co} S'(l', L')} \quad \frac{P_1 \xrightarrow{(\alpha^-, w)}_{co} P'_1 \quad \alpha \notin \mathcal{L}}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha^-, w)}_{co} P'_1 \boxtimes_{\mathcal{L}} P_2} \\
\\
\frac{P_1 \xrightarrow{(\alpha^-, w_1)}_{co} P'_1 \quad P_2 \xrightarrow{(\alpha^-, w_2)}_{co} P'_2 \quad \alpha \in \mathcal{L}}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha^-, w_1 \oplus w_2)}_{co} P'_1 \boxtimes_{\mathcal{L}} P'_2} \quad \frac{P_2 \xrightarrow{(\alpha^-, w)}_{co} P'_2 \quad \alpha \notin \mathcal{L}}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha^-, w)}_{co} P_1 \boxtimes_{\mathcal{L}} P'_2}
\end{array}$$

Fig. 4. The completion relation $\rightarrow_{co} \subseteq \mathcal{C} \times \Theta^- \times \mathcal{C}$.

The completion relation is $\rightarrow_{co} \subseteq \mathcal{C} \times \Theta^- \times \mathcal{C}$ where Θ^- is an adaptation of Θ with labels of the form (α^-, w) and w is defined as for the start relation we discussed. The completion relation is defined as the minimal relation satisfying the rules of Fig. 4.

Formally, for a species $S(l, L)$ it is possible to get the instance of a currently running action α , if any, by applying function *pick*. More precisely, this permits us to get, from all the possible instances of actions α , the first which has been scheduled, *pick* αL , and, hence, the first which will terminate. If the species is involved as a product, then it is necessary to increase, as defined by the delay-as-duration approach, its concentration level by adding the scheduled products. Otherwise, whatever the role of the species, its concentration level must remain constant. Notice that the reactants were already removed at the derivation of a start relation. Independently of the role of the species in the action, the scheduling list is modified by means of the function *del*, hence a new list *del* αL is produced by removing from L the entry which was computed by function *pick*.

It is unnecessary to state constraints for the completion of a currently running action, as the bounds on the levels will have been checked when the action started.

The completion of the action α , in the style of the ST semantics, is denoted by the action symbol α^- , exhibited as a label for all the completion derivations. The other part of the label, namely the list w , is defined like the one exhibited by the start relation. The treatment of composition of this relation with the other operators is straightforward and very similar to the way compositions are dealt with in the derivations of the start relation.

Some further considerations are worth noting. Firstly, this relation is a new one with respect to the Bio-PEPA semantics. When actions have no delays we can assume that the start of an exponential distributed action and its completion coincide, and hence this relation would be redundant since all relevant information can be derived from the starting of the action. When a deterministic delay is associated with an action the role of this relation is to model the completion of an action. To do so it chooses actions to terminate from the list which is associated with the species, namely the list of actions currently running. The start relation, differently, chooses the action to fire from the species definition.

Furthermore, as we want the completion relation to exhibit quantitative information in order to recompute the functional rate of the action at the moment at which it started, then the labels exhibited by this relation are very similar to those exhibited by the start relation, even if they are computed starting from *pick* αL . This permits us to have a unique policy for computing the functional rates from the input lists, obtained by derivations of the transitions of these relations.

The stochastic relation

The stochastic relation permits us to associate rates with transitions. Also, this transition permits us to observe changes in a Bio-PEPA system due to either the starting or the completion of an action. The stochastic relation is $\rightarrow_s \subseteq \overline{\mathcal{P}} \times \overline{\Gamma} \times \overline{\mathcal{P}}$ where

$$\overline{\Gamma} = \{(\alpha^*, r_\alpha, \sigma_\alpha) \mid \alpha \in \mathcal{A}, * \in \{+, -\}, r_\alpha, \sigma_\alpha \in \mathbb{R}^+\}.$$

As this relation is defined on the set $\overline{\mathcal{P}}$, namely the set of all possible Bio-PEPA systems with process configurations, whenever we refer to the semantics of a system $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, P \rangle$, where P is a Bio-PEPA process, we assume we apply the stochastic relation to the system $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, \mu(P) \rangle$. Again, this is necessary because P is not a process configuration, and we want to build, from P , the corresponding initial configuration $\mu(P)$, and then we want to apply the semantics to the system.

The stochastic relation is defined as the minimal relation satisfying the rules given in Fig. 5. Formally, the starting of an action α , obtained by composition with a derivation of the start relation, is denoted by symbol α^+ . The completion of an action is obtained by composition with a derivation of the completion relation, as denoted by symbol α^- . The rates of actions are computed as in Bio-PEPA, namely as $r_\alpha = f_\alpha[w, \mathcal{N}, \mathcal{K}]h^{-1}$. As in Bio-PEPA, r_α represents the parameter of an exponential distribution and, as expected, all activities enabled attempt to proceed but only the fastest succeeds. For the explanation of how the rates are computed because of the levels we refer to [14]. For any possible derivation of the stochastic relation,

$$\begin{array}{c}
P \xrightarrow{(\alpha^+, w)}_{st} P' \quad r_\alpha = f_\alpha[w, \mathcal{N}, \mathcal{K}]h^{-1} \\
\hline
\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, P \rangle \xrightarrow{(\alpha^+, r_\alpha, \sigma(\alpha))}_s \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, P' \rangle \\
\\
P \xrightarrow{(\alpha^-, w)}_{co} P' \quad r_\alpha = f_\alpha[w, \mathcal{N}, \mathcal{K}]h^{-1} \\
\hline
\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, P \rangle \xrightarrow{(\alpha^-, r_\alpha, \sigma(\alpha))}_s \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, P' \rangle
\end{array}$$

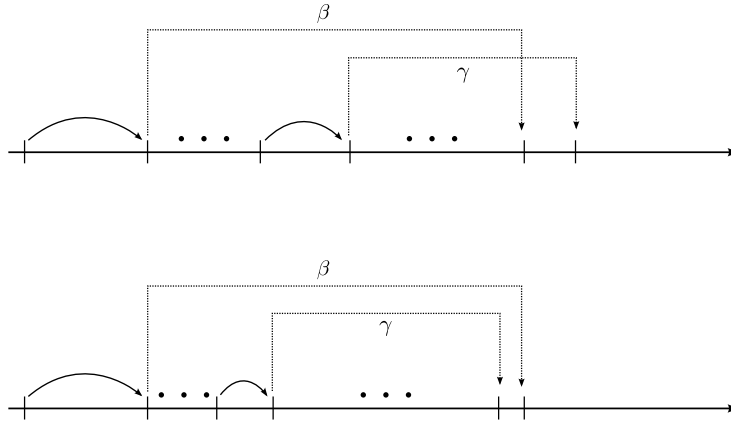
Fig. 5. The stochastic relation $\rightarrow_s \subseteq \overline{\mathcal{P}} \times \overline{\Gamma} \times \overline{\mathcal{P}}$.

Fig. 6. Timing aspects in Bio-PEPAD.

the value $\sigma(\alpha)$ denotes the delay of the action α . Thus the labels $(r_\alpha, \sigma_\alpha)$ denotes a random variable $(X + Y)$ where X is exponential with mean $1/r_\alpha$ and Y is deterministic with parameter σ_α .

A SLTS can be defined for a Bio-PEPA system with delays.

Definition. The *Stochastic Labeled Transition Systems* (SLTS) for a Bio-PEPAD system is $(\overline{\mathcal{P}}, \overline{\Gamma}, \rightarrow_s)$ where \rightarrow_s is the minimal relation satisfying the rules given in Fig. 5.

Before discussing analysis techniques for Bio-PEPAD systems, timing aspects in the semantics we defined are worth discussing.

On timing aspects in Bio-PEPAD. As in most stochastic process algebras the underlying SLTS does not contain an explicit quantitative notion of time. In non-delayed systems this is in perfect agreement with CTMCs since time can be retrieved by using distributions, whereas in systems with delays the correspondence with non-Markov processes needs to be established. By means of the ST semantics, in Bio-PEPAD a qualitative notion of time can be retrieved by observing state changes induced by either the start or the completion of an action. Moreover, by construction instances of an action complete while respecting their starting order.

Start and completion relations are not aware of kinetic information appearing in systems, hence they generate all the possible behaviors for a process configuration. Among these there are some which, in view of the kinetic information of the system, are not physically possible despite their appearance in the defined SLTS. Note that this may make model-checking of Bio-PEPAD systems even less tractable than Bio-PEPA systems since the SLTS of a system with delays is more complex than the corresponding non-delayed SLTS, as will be discussed in Section 5.

We expand this point with an informal example. Let us assume a $P \in \mathcal{C}$ where two actions β and γ are started in that order. The completion relation allows us to derive the completion of both actions as $P \xrightarrow{(\beta^-, w_1)}_{co} P'$ and $P \xrightarrow{(\gamma^-, w_2)}_{co} P''$, and by construction this instance of β is the first to complete among all β actions, and similarly for γ . However, no property on the order of completion between β and γ can be stated. In the case $\sigma(\beta) > \sigma(\gamma)$ both the derivations are correct: deriving P' means that γ started and completed during the completion of β ; in the other case γ completed before the completion of β , and this is possible since $\sigma(\beta) > \sigma(\gamma)$. Both cases are shown in Fig. 6.

However, if γ started before β the only correct derivation models the completion of γ , since there is no chance that β completes before γ when $\sigma(\beta) > \sigma(\gamma)$. Hence one of the two derivations should be discarded.

By generalizing this example, an action β should complete if all the actions starting after β have longer duration. The stochastic relation is the only one capable of discovering this, hence it could filter out inappropriate cases. The stochastic relation we defined is not precise in this sense. However, we have at least two major reasons supporting our choice.

The first is a probabilistic motivation: to any execution of the non-Markovian stochastic processes logically connected to this SLTS such transitions have probability 0, which pragmatically makes them absent from a probabilistic perspective. The second motivation is more technical. Intuitively, compositionality requires us to have scheduling lists local to each species

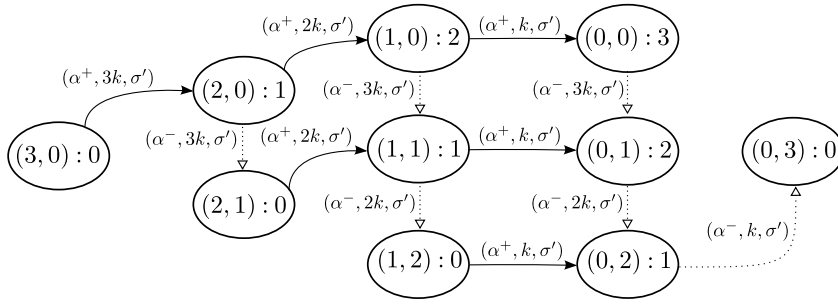


Fig. 7. The SLTS for the Bio-PEPAD example.

making it impossible to establish a global ordering relation for completion. In fact, in a semantics without explicit time the only ordering we have is given by the positions in the scheduling lists.

A Bio-PEPAD toy example

In order to clarify the modeling with Bio-PEPAD we present a simple extension of the Bio-PEPA toy model discussed above. In order to switch to the Bio-PEPAD framework we assume that the reaction $A \xrightarrow{k} B$, denoting the transformation of an element of species A into an element of species B with a kinetic constant k , is now enriched with a delay $\sigma' > 0$, giving rise to the definition of the reaction $A \xrightarrow{k, \sigma'} B$. We assume the initial state described by the vector $\mathbf{x}_0 = (3, 0)^T$.

Since we defined a conservative extension of Bio-PEPA, we are able to fully reuse the Bio-PEPA specification for this model, namely the process definitions $A \stackrel{\text{def}}{=} (\alpha, 1) \downarrow A$, $B \stackrel{\text{def}}{=} (\alpha, 1) \uparrow B$, and $A \bowtie_{\{\alpha\}} B$. Also, the kinetic information about the system is preserved, namely $f_\alpha = f_{MA}(k')$ when $k' = k$. Conversely, the information about the delay of α , which is not present in Bio-PEPA, is defined according to the function $\sigma(\alpha) = \sigma'$.

By considering the same Bio-PEPA levels, the initial configuration of the process, obtained by applying function μ , is the following

$$A(3, []) \bowtie_{\{\alpha\}} B(0, []).$$

By applying the stochastic relation to the system with this process configuration we obtain all the possible evolutions of the configuration. The obtained SLTS, as expected, is finite, and, because of the delays, it corresponds to a non-Markovian stochastic process. Intuitively, there is a one-to-one correspondence between both the states and the transitions of the SLTS and those of the stochastic process, analogous to the relation between the SLTS of a Bio-PEPA system and the underlying CTMC.

A graphical representation of the state-transitions for the process is given in Fig. 7. In that figure, all the states are represented as circles where the notation $(n_1, n_2) : m$ represents the discrete levels of concentration of the species A , n_1 , and B , n_2 . The number m represents the number of instances of the unique possible action α currently scheduled in the state. Note that $n_1 + n_2 + m = 3$. All the arrows represent stochastic derivations of the whole system, where the labels are exactly those computed by that relation. The full arrows represent stochastic derivations based on start derivation, empty arrows represent stochastic derivations based on completion derivation. For this particular example, any empty arrow built from a derivation with a rate r refers to the completion of the unique action started with the same rate r .

Fig. 8 presents a table showing the explicit mapping of the states described in Fig. 7 and the corresponding process configuration obtained by the semantics in the SLTS. For the sake of clarity, as in this simple example there is just one action, α , and A always participates in that action as a reactant and B as a product, this information is omitted from the scheduling lists. Note that once an action completes the states of the SLTS contain the necessary information to re-compute the rate at which the action started.

As expected, this system, starting from the initial configuration, namely state $(3, 0) : 0$, eventually reaches the final state $(0, 3) : 0$, which corresponds to the final configuration $A(0, []) \bowtie_{\{\alpha\}} B(3, [])$ and to the vector $(0, 3)^T$.

4. Analysis techniques

In this section we present some analysis techniques for Bio-PEPAD systems analogous to those presented in [14] for Bio-PEPA systems. Firstly, we present the automatic translation of a Bio-PEPAD system into a set of Delay Differential Equations (DDEs). Secondly, we discuss how to apply a Delay Stochastic Simulation Algorithm (DSSA) to compute the stochastic time-evolution of a Bio-PEPAD model. Thirdly, we discuss the encoding of Bio-PEPAD processes in Generalized Semi-Markov Processes (GSMPs).

state	process configuration
(3, 0) : 0	$A(3, [\]) \bowtie_{\{\alpha\}} B(0, [\])$
(2, 0) : 1	$A(3, [(3, 1)]) \bowtie_{\{\alpha\}} B(0, [(0, 1)])$
(2, 1) : 0	$A(2, [\]) \bowtie_{\{\alpha\}} B(1, [\])$
(1, 0) : 2	$A(1, [(3, 1), (2, 1)]) \bowtie_{\{\alpha\}} B(0, [(0, 1), (0, 1)])$
(1, 1) : 1	$A(1, [(2, 1)]) \bowtie_{\{\alpha\}} B(1, [(1, 1)])$
(1, 2) : 0	$A(1, [\]) \bowtie_{\{\alpha\}} B(2, [\])$
(0, 0) : 3	$A(0, [(3, 1), (2, 1), (1, 1)]) \bowtie_{\{\alpha\}} B(0, [(0, 1), (0, 1), (0, 1)])$
(0, 1) : 2	$A(0, [(2, 1), (1, 1)]) \bowtie_{\{\alpha\}} B(1, [(1, 1), (1, 1)])$
(0, 2) : 1	$A(0, [(1, 1)]) \bowtie_{\{\alpha\}} B(2, [(2, 1)])$
(0, 3) : 0	$A(0, [\]) \bowtie_{\{\alpha\}} B(3, [\])$

Fig. 8. A table stating the correspondence between the states represented in Fig. 7 and the process configurations obtained by the semantics.

4.1. Translation in Delay Differential Equations

Whenever phenomena presenting a delayed effect are described by differential equations, we move from ODEs to DDEs. In DDEs the derivatives at the current timepoint depend on some past states of the system. The simplest form of DDE considers constant delays $\sigma_1 > \dots > \sigma_n \geq 0$ and consists of an equation of the form

$$\frac{dX}{dt} = \varphi_X(t, \{X(t - \sigma_i) \mid i = 1, \dots, n\})$$

where $X(t - \sigma_i)$ denotes the state of the system at the past timepoint $t - \sigma_i$. This form of DDE allows models to describe events which have a fixed duration. Hence it is natural, in the context of Bio-PEPAD, to reason about the translation of a model into a set of DDEs. Furthermore, similar work has been presented in [14] for translating a Bio-PEPA system into a set of ODEs.

In order to define the encoding it is important to recall that we defined Bio-PEPAD in terms of Bio-PEPA. This means that, given a system specification $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, P \rangle$ where P is a valid Bio-PEPA process, we just need to modify the algorithm defined in [14] to add the information provided by σ concerning the delays. Formally, the results for Bio-PEPA permit us to encode $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, P \rangle$ in a set of ODEs by using the definition of the stoichiometry matrix associated with P .

The algorithm presented in [14] consists of three steps. In the first, by the syntactic definition of the components, the stoichiometry matrix $D = \{d_{i,j}\}$ is defined, in the second the kinetic law vector $\overline{v_{KL}}$ is derived and in step three the deterministic variables are associated with the components.

We discuss the steps of the algorithm:

- (1) As in Bio-PEPA, the stoichiometry matrix is $D \in \mathbb{N}^{n \times m}$ if the system contains n distinct species which can perform m actions. Here, we assume we enumerate the actions as $\alpha_1, \dots, \alpha_m$ and the species in the system as S_1, \dots, S_n . The entry $d_{i,j}$, representing the change in the levels induced by performing action α_j with species S_i , is defined as follows:

$$d_{i,j} = \begin{cases} -\kappa_{i,j} & \text{if } (\alpha_j, \kappa_{i,j}) \downarrow \text{ is an action for species } S_i \\ +\kappa_{i,j} & \text{if } (\alpha_j, \kappa_{i,j}) \uparrow \text{ is an action for species } S_i \\ 0 & \text{otherwise.} \end{cases}$$

- (2) The definition of the Bio-PEPAD m -dimensional kinetic law vector $\overline{v_{KL}}$ is different from Bio-PEPA. In this step, we build, instead of a set of ODEs, a set of DDEs. We formally define the entries in the vector only for the well-known kinetic functions (f_{MA} , f_{MM} and f_H), arbitrary functions must be appropriately encoded by the modeler. Here we denote with S_1 and S_2 either two species involved as reactants in a mass-action kinetic, or two species involved as an enzyme S_1 and a substrate S_2 , respectively, in a Michaelis–Menten kinetic, or, in the case of Hill kinetics, the only reactant is from species S_1 . With x_i we denote the deterministic variable representing species S_i . The entry $\overline{v_{KL_i}}$ of the vector is defined as follows:

$$\overline{v_{KL_i}} = \begin{cases} kx_1(t - \sigma(\alpha_i))x_2(t - \sigma(\alpha_i)) & \text{if } f_{\alpha_i} = f_{MA}(k) \\ \frac{vx_1(t - \sigma(\alpha_i))x_2(t - \sigma(\alpha_i))}{K + x_2(t - \sigma(\alpha_i))} & \text{if } f_{\alpha_i} = f_{MM}(v, K) \\ \frac{vx_1(t - \sigma(\alpha_i))^p}{K + x_1(t - \sigma(\alpha_i))^p} & \text{if } f_{\alpha_i} = f_H(v, K, p). \end{cases}$$

- (3) As in Bio-PEPA, now we associate the variable x_i with each species component S_i and so define the n -dimensional vector \bar{x} .

The DDE system can be defined in the same way as the ODE system in Bio-PEPA, namely as

$$d\bar{x}/dt = D \bar{v}_{KL}$$

where \bar{x} and D are the results of step (3) and (1) of the algorithm, respectively. The initial conditions are, however, different from the ones defined for ODEs. In particular, the DDEs, because of the delays, must be defined also in the interval $[t_0 - \sigma(\alpha); t_0]$ where α is the action with maximum delay.

It is not possible to define a universal initial condition for the DDEs systems as every possible configuration will affect the dynamics of the whole system. Sometimes the initial conditions of a species S are defined via a constant function $\varphi_S(t)$ for $t \in [t_0 - \sigma(\alpha); t_0]$ such that $\varphi_S(t) = h l_{S,0}$ where $l_{S,0}$ is the initial concentration level for S in the Bio-PEPAD model and h is the step size for the concentration levels. In general, we leave this part of the translation to the modeler who will tune the initial conditions with respect to the specification of the target system.

Mapping a 2-reaction Bio-PEPAD model

Let us extend the Bio-PEPAD model presented previously to consider a reversible transformation, i.e.



This pair of reactions can be modeled by simply extending the single-reaction model. Indeed, we define the processes

$$A \stackrel{\text{def}}{=} (\alpha, 1)\downarrow + (\beta, 1)\uparrow \quad B \stackrel{\text{def}}{=} (\alpha, 1)\uparrow + (\beta, 1)\downarrow \quad A(3) \boxtimes_{\{\alpha, \beta\}} B(0)$$

where β models the new reaction. Encoding this simple process is straightforward. Firstly, the definition of the stoichiometry matrix is

$$D = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

since on the columns we assume to have the reactions in order of appearance, and on the rows the species A and B . As expected, $d_{1,1} = -1$ reflects the transformation of one level of A and $d_{1,2} = 1$ the production of one level of B .

The second step defines the kinetic law vector. Since there are two reactions, the vector is 2-dimensional, and is defined as

$$\bar{v}_{KL} = \begin{pmatrix} kx_1(t - \sigma(\alpha)) \\ k'x_2(t - \sigma(\beta)) \end{pmatrix} = \begin{pmatrix} kx_1(t - \sigma') \\ k'x_2(t - \sigma'') \end{pmatrix}.$$

The third step associates the names x_1 and x_2 with the species A and B , respectively. Finally, the DDE system $d\bar{x}/dt = D \bar{v}_{KL}$ is the following:

$$\frac{dx_1}{dt} = -kx_1(t - \sigma') + k'x_2(t - \sigma'') \quad \frac{dx_2}{dt} = kx_1(t - \sigma') - k'x_2(t - \sigma'').$$

By defining some initial conditions in $[t_0 - \max\{\sigma', \sigma''\}; t_0]$ the system could be either analytically or numerically studied.

4.2. Stochastic simulation of Bio-PEPAD systems

The stochastic simulation of biological systems is typically based on the SSA by Gillespie [24] and its variants. However, neither the SSA, nor its variants designed only for Markovian actions, are able to deal with actions with delays. As a consequence, some Delay Stochastic Simulation Algorithms [5,1,12] (DSSAs) have been defined to perform stochastic simulation of systems where actions have a fixed delay following the delay-as-duration approach. In [12] it has been shown that these DSSAs produce a single time-trajectory of the underlying Generalized Semi-Markov Process.

In this section we briefly explain how to perform the stochastic simulation of a Bio-PEPAD system by using the DSSAs presented in [1,5], where all the reactions follow a delay-as-duration approach and whose definition is recalled as Algorithm 1. As above the Bio-PEPA approach forms the basis for Bio-PEPAD analysis. In particular, we are able to re-use parts of the method defined in [14] to perform stochastic simulation of Bio-PEPA systems using the SSA.

The main steps in preparing a Bio-PEPAD system for the application of the DSSA are two: define the algebraic representation of a process, and create the reactions to simulate. We introduce a family of functions

$$\{(\llbracket _ \rrbracket)_n : \mathcal{C} \cup \mathcal{P} \rightarrow \mathbb{N}^n \mid n \in \mathbb{N}\}$$

for the encoding of either a Bio-PEPA process or a Bio-PEPAD process configuration in an n -dimensional vector such that:

$$\begin{aligned} (\llbracket S_1(l_1) \boxtimes_{\mathcal{L}_1} \dots \boxtimes_{\mathcal{L}_m} S_m(l_m) \rrbracket)_m &= (l_1, \dots, l_m)^T \\ (\llbracket S_1(l_1, L_1) \boxtimes_{\mathcal{L}_1} \dots \boxtimes_{\mathcal{L}_n} S_n(l_n, L_n) \rrbracket)_n &= (l_1, \dots, l_n)^T. \end{aligned}$$

Algorithm 1 DSSA DDA(t_0, \mathbf{x}_0, T)

```

1:  $t \leftarrow t_0; \mathbf{x} \leftarrow \mathbf{x}_0; S \leftarrow \emptyset;$ 
2: while  $t < T$  do
3:    $a_0(\mathbf{x}) \leftarrow \sum_{j=1}^M a_j(\mathbf{x});$ 
4:   let  $r_1, r_2 \sim U[0, 1];$ 
5:    $\tau \leftarrow a_0(\mathbf{x})^{-1} \ln(r_1^{-1});$ 
6:   let  $S_{t,\tau} = \{(t'', v'') \in S \mid t'' \in (t, t + \tau]\};$ 
7:   if  $S_{t,\tau} \neq \emptyset$  then
8:      $(t', v') \leftarrow \min\{S_{t,\tau}\};$ 
9:      $\mathbf{x} \leftarrow \mathbf{x} + v'; t \leftarrow t'; S \leftarrow S \setminus \{(t', v')\};$ 
10:  else
11:    let  $j$  such that  $\sum_{i=1}^{j-1} a_i(\mathbf{x}) < r_2 \cdot a_0(\mathbf{x}) \leq \sum_{i=1}^j a_i(\mathbf{x});$ 
12:     $\mathbf{x} \leftarrow \mathbf{x} + v_j^r; t \leftarrow t + \tau; S \leftarrow S \cup \{(t + \tau + \sigma_j, v_j^p)\};$ 
13:  end if
14: end while

```

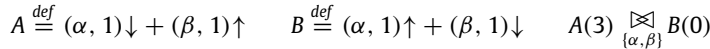
Since we assume we have a well-defined Bio-PEPA model all the species S_i are distinct, and are represented by a distinct element in the resulting vector. In the following, we use $\langle _ \rangle$ to denote the function mapping a Bio-PEPA process or a Bio-PEPAD process configuration to an appropriate vector-space. For instance, for the example we previously discussed, the encoding is such that $\langle A(3, [\]) \bowtie_{\{\alpha\}} B(0, [\]) \rangle = (3, 0)^T$. Given an initial system $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, P \rangle$ and $\mathbf{x}'_0 = \langle P \rangle$ a vector \mathbf{x}_0 describing the initial number of molecules to be simulated is defined by $\mathbf{x}_0[i] = \mathbf{x}'_0[i] \times h \times N_A \times v$ where h is the step size of the system, v is the volume of the target compartment and N_A is the Avogadro number.

Secondly, the actual rates of the reactions have to be defined, case by case, using the parameters in \mathcal{F} . Since \mathcal{F} is defined in the same way for both Bio-PEPA and Bio-PEPAD, the techniques developed in [14] to derive rates from \mathcal{F} the actual rates can be also applied in the context of Bio-PEPAD. Once these two steps have been performed, the resulting system can be simulated by the DSSA where all the actions follow a delay-as-duration approach, given as Algorithm 1.

The algorithm works as follows: in state \mathbf{x} at time t , after the propensity functions (denoted as a_j , [24]) have been evaluated as $a_j(\mathbf{x})$, the putative time for next reaction τ is calculated. If there are actions completing in $[t, t + \tau]$, and this is discovered by using a set-representation of the scheduling list (denoted as S), then τ is discarded and the first reaction to complete (obtained evaluating $\min\{S_{t,\tau}\}$) is fired with the delay-as-duration approach. The notation v_j^p represents the stoichiometry vector for the products of reaction R_j which are, in this approach, the ones which need to be added to \mathbf{x} to complete the firing of the scheduled reaction. Notice that such a vector is scheduled at step (12) and eventually used at step (9). Conversely, if no actions will complete before $t + \tau$ ($S_{t,\tau} = \emptyset$), the system can remove the reactants for the next reaction to fire by performing $\mathbf{x} = \mathbf{x} + v_j^r$, where v_j^r represents the stoichiometry vector for the reactants of reaction R_j , and schedule the insertion of the products at time $t + \sigma_j + \tau$, if σ_j is the delay of the reaction. Notice that steps (12) and (9) are modeled by the Bio-PEPA start and completion relations, respectively. For more detailed considerations about this DSSA, as well as its proof of correctness and the derivation of the *Delay Chemical Master Equation* underlying the simulated systems we refer to [12].

Mapping the 2-reaction Bio-PEPAD model

Let us consider the same Bio-PEPAD model translated in DDEs in the previous section. By applying the techniques we have introduced the processes



can be encoded in the 2-reactions model



and the initial state vector can be defined as

$$\mathbf{X}(t_0) = \begin{pmatrix} n_0^A \times h \times N_A \times v \\ n_0^B \times h \times N_A \times v \end{pmatrix} = \begin{pmatrix} 3 \times h \times N_A \times v \\ 0 \end{pmatrix}.$$

Finally, $t_0, \mathbf{X}(t_0)$ and $\{R_1, R_2\}$, together with the information about the propensity functions for the reactions, can be used as input for the DSSA to analyze the model.

4.3. Bio-PEPAD processes as Generalized Semi-Markov Processes

Since Algorithm 1 provides a single time-trajectory of a Generalized Semi-Markov Process (GSMP) [19,26,10] which underlies a stochastic model with delays following the delay-as-duration approach [12], in this section we show how to build a GSMP from a Bio-PEPAD system.

We recall the definition of finite-state GSMPs as in [19]. Let $\mathbf{E} = \{e_1, \dots, e_n\}$ be a finite set of *events*. For any state $s \in S$, let $s \mapsto E(s)$ be a mapping from s to a non-empty subset of \mathbf{E} denoting the *active events* in state s . When in state s the occurrence of one or more events triggers a state transition, the next state s' is chosen according to a probability distribution $p(s'; s, E^*)$ where $E^* \subseteq E(s)$ is the set of active events which are triggering the state transition. Clocks are associated with events and, in state s , the clock associated with event e decays at rate $r(s, e)$. In our case, and in most applications, the rate of decay of clocks is always 1. When, in a state s , there are no outgoing transitions, i.e. $E(s) = \emptyset$, the state s is said to be *absorbing* and it models a terminating process. The set of possible *clock-reading vectors* when the state is s is

$$C(s) = \{c = (c_1, \dots, c_M) \mid c_i \in [0, \infty) \wedge c_i > 0 \Leftrightarrow e_i \in E(s)\}$$

where c_i is the value of the clock associated with e_i ; $c_i \in \mathcal{C}_\ell$ where \mathcal{C}_ℓ is the set of clock evaluations. In state s with clock-reading vector c , the time to the next transition is

$$t^*(s, c) = \min_{\{i \mid e_i \in E(s)\}} c_i / r(s, e_i)$$

where $c_i / r(s, e_i) = +\infty$ when $r(s, e_i) = 0$. The set of events triggering the state transition is then

$$E^*(s, c) = \{e_i \in E(s) \mid c_i - t^*(s, c)r(s, e_i) = 0\}.$$

When a state transition from s to s' is triggered the events E^* expire, leaving $E'(s) = E(s) \setminus E^*$. Moreover some new events are created; this set of *new events* is $E(s') \setminus E'(s)$. For these events e' a clock value x is generated by a *distribution-assignment* function $F(x; s', e', s, E^*)$ such that $F(0; s', e', s, E^*) = 0$ and $\lim_{x \rightarrow \infty} F(x; s', e', s, E^*) = 1$. For the *old events* in $E(s') \cap E'(s)$ the clock value in state s at the time when the transition was triggered is maintained in s' . In s' events in $E'(s) \setminus E(s')$ are canceled and the corresponding clock value is discarded. The GSMP is a continuous-time stochastic process $\{X(t) \mid t \geq 0\}$ recording the state of the system as it evolves and its semantics is given in terms of a general state space Markov chain storing both the state of the process and the clock-reading vectors [26]. We can summarize the definition of a GSMP as follows [19].

Definition. $(S, s_0, \mathbf{E}, e_0, E, C, N, F)$ is a *Generalized Semi-Markov Process* (GSMP) with S , a non-empty finite set of *states* and $s_0 \in Z$ as initial state, \mathbf{E} as the non-empty set of *events* with $e_0 \in \mathbf{E}$, $E : S \rightarrow \wp_{fin}(\mathbf{E})$ the *event-assignment* function and a unique initial event $E(s_0) = \{e_0\}$, $C : \mathbf{E} \rightarrow \mathcal{C}_\ell$ the *clock-assignment* function, $N : S \times \wp_{fin}(\mathbf{E}) \rightarrow (S \rightarrow [0, 1])$ the *next-state* function and $F : \mathcal{C}_\ell \rightarrow (\mathbb{R} \rightarrow [0, 1])$ the *distribution-assignment* function, such that $F(x)(0) = 0$. The set of all possible GSMPs is denoted as \mathcal{G} .

Notice that we restrict to the case of a unique initial state since, in Bio-PEPAd, we have a unique initial state, the one in which no actions are running. We can now investigate the relationship between Bio-PEPAd systems and GSMPs. However, we first restrict our attention to the set \mathcal{B} of Bio-PEPAd systems satisfying both the following assumptions:

- (i) systems in \mathcal{B} have a finite state-space;
- (ii) systems in \mathcal{B} are such that, whenever an action starts, the concentration level in at least one species component changes.

Assumption (i) is merely technical, since we will relate Bio-PEPAd systems with finite-state GSMPs. Assumption (ii), which draws a clear connection with GSMPs as we will discuss later, is actually reasonable in the context of biochemical systems or, more generally, biological systems. In fact, it means that, for any action appearing in the process, there exists at least one species which has a reactant-prefix transition with stoichiometry greater than 0. Chemically, this means that reactions which do not use reactants are not allowed. We point out that, in a delayed framework, the notions of *usage* and *consumption* of reactants are not necessarily the same, whereas in the non-delayed framework they coincide. So, for instance, a reaction $A \xrightarrow{k} A + B$ in the non-delayed framework does not consume molecule A . In contrast, under the delay-as-duration approach it uses a molecule A , in the sense that the molecule is removed from the state of the simulation at the time of the start of the reaction but, since the molecule is re-inserted in the state at the completion of the reaction, then the molecule is not consumed. The form of reaction which we disallow would be $\emptyset \xrightarrow{k} A$.

We now establish the relationship between Bio-PEPAd systems and GSMPs. Recall that a Bio-PEPAd system is $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, \mu(P) \rangle$. For convenience and to shorten the notation we will denote this by $\langle \mathcal{T}, \sigma, P_C \rangle$ where $\mu(P) = P_C$. According to Definition 5 the semantics of this configuration is the SLTS rooted in $\langle \mathcal{T}, \sigma, P_C \rangle$. We let S_P denote the set of states in this SLTS and $Act(P_C)$ the set of action types which appear in it. For an arbitrary state, $P_i \in S_P$, we write $P_i \xrightarrow{\ell}_s P_j$ to denote the transition in the SLTS from P_i to P_j exhibiting label ℓ ; we remark that each such element of the stochastic relation is derived from either the start or the completion of an action. Moreover, each action start is governed by an exponential distribution whereas each completion is deterministically timed. For any process P we let $\mathcal{A}^+(P)$ denote those actions which can start in P and $\mathcal{A}^-(P)$ those actions which may complete in P , i.e.

$$\mathcal{A}^+(P) = \{\alpha \mid \exists P', w. P \xrightarrow{(\alpha^+, w)}_{st} P'\}$$

$$\mathcal{A}^-(P) = \{\alpha \mid \exists P', w. P \xrightarrow{(\alpha^-, w)}_{co} P'\}.$$

We also recall that $\langle P \rangle$ associates a discrete-state vector with the process configuration P .

Definition. For any Bio-PEPAD system $\langle \mathcal{T}, \sigma, P_C \rangle \in \mathcal{B}$ we define $G_P(\langle \mathcal{T}, \sigma, P_C \rangle) = (S, s_0, \mathbf{E}, e_0, E, C, N, F)$ where:

- $S = \{\langle P_i \rangle \mid P_i \in S_P\}$;
- $s_0 = \langle P_C \rangle$;
- $\mathbf{E} = \mathbf{E}^+ \cup \mathbf{E}^-$ where $\mathbf{E}^+ = \{e_i \mid s_i \in S\}$ are exponentially-timed events and $\mathbf{E}^- = \{e_{\alpha,j} \mid \alpha \in \mathcal{A}^+(P_j) \wedge P_j \in S_P\}$ are deterministically-timed events;
- $E : S \rightarrow \wp_{fin}(\mathbf{E})$ where, for any $P_i \in S_P$, E is defined as

$$E(\langle P_i \rangle) = \{e_i\} \cup \{e_{\alpha,j} \mid \alpha \in \mathcal{A}^+(P_j) \wedge \alpha \in \mathcal{A}^-(P_i) \wedge P_j \in S_P\};$$

- $C : \mathbf{E} \rightarrow \mathcal{C}_\ell$ where, for any $e \in \mathbf{E}$, C is defined as

$$C(e) = \begin{cases} c_i = \exp\left(\sum_{\alpha \in \mathcal{A}^+(P_i)} r_\alpha\right) & \text{if } e \equiv e_i \\ c_{\alpha,j} = \sigma(\alpha) & \text{if } e \equiv e_{\alpha,j}; \end{cases}$$

- $N : S \times \wp_{fin}(\mathbf{E}) \rightarrow (S \rightarrow [0, 1])$ where, for any $P_i \in S_P$ and $E^* \subset \mathbf{E}$, N is defined as

$$N(\langle P_i \rangle, E^*) = \begin{cases} \langle P_j \rangle \longrightarrow r_\alpha / \sum_{\beta \in \mathcal{A}^+(P_i)} r_\beta & \text{if } E^* \subset \mathbf{E}^+ \wedge \langle \sigma, P_i \rangle \xrightarrow{\alpha^+, r_\alpha}_s \langle \sigma, P_j \rangle \\ \left(\langle P_i \rangle + \sum_{e_{\alpha,k} \in E^*} v_\alpha^p \right) \rightarrow 1 & \text{if } E^* \subseteq \mathbf{E}^-; \end{cases}$$

with v_α^p the update vector associated with the completion of action α , when it started in process P_k ;

- $F : \mathcal{C}_\ell \rightarrow (\mathbb{R} \rightarrow [0, 1])$ defined as

$$F(c'; s', e', s, E^*) = \begin{cases} 1 - \exp\left(x \sum_{\alpha \in \mathcal{A}^+(P')} r_\alpha\right) & \text{if } c' \equiv c_i \wedge s' = \langle P' \rangle \\ H(x - \sigma(\alpha)) & \text{if } c' \equiv c_{\alpha,j}, \end{cases}$$

where $H(\cdot)$ is the Heaviside function which is the unit step function with value 0 for negative arguments and value 1 for positive ones.

It remains to show that $G_P(\langle \mathcal{T}, \sigma, P_C \rangle)$ is a GSMP, which we do in the following theorem.

Theorem 4.1. For any Bio-PEPAD system $\langle \mathcal{T}, \sigma, P_C \rangle \in \mathcal{B}$ it holds that $G_P(\langle \mathcal{T}, \sigma, P_C \rangle)$ is a GSMP, namely $G_P : \mathcal{B} \rightarrow \mathcal{G}$.

Proof. By definition S is a non-empty set since $|S| = |S_P|$ and S_P must contain at least P_C . Moreover, by assumption (i) for all Bio-PEPAD systems in \mathcal{B} , S_P is finite so it follows that S is finite. Furthermore, it follows immediately that $\langle P_C \rangle$ is the initial state.

The set \mathbf{E} is comprised of two subsets. \mathbf{E}^+ , the exponentially-timed events model the start of a new action. There is one such event for each state of the system, whose rate is the minimum of the exponential delays associated with all the possible new actions in that state. Let s_i be an arbitrary state of S , corresponding to $P_i \in S_P$. Then let e_i be the exponentially-timed event with rate $\sum_{\alpha \in \mathcal{A}^+(P_i)} r_\alpha$. \mathbf{E}^- , the deterministic-timed events model the completion of an action which is already running. There is one such possible action for each action type, and each starting state. Thus $|\mathbf{E}| = |S|(1 + |\text{Act}(P_C)|)$ but note that this is an upper bound on the possible number of events in the system, since, depending on the structure of the system, some of these events may not be reachable. Since in the initial state of the Bio-PEPA system $\mu(P)$ there are no actions running, there is a unique event in this state representing the start of an action, and this is the event e_0 . In the Bio-PEPAD system time advances globally at a constant rate so each event $e \in \mathbf{E}$ decays with rate $r(e, s_i) = 1$ for any state $s \in S$.

The function E maps states to subsets of events. These are precisely those events corresponding to actions which may start or complete in the corresponding process configuration. All action starts are represented by the single exponentially-timed event for that state.¹ In contrast each action completion is represented as a distinct event and there is one event for each action running in the current state. Notice that assumption (ii) implies that $\langle P_i \rangle \neq \langle P_j \rangle$ since the two vectors differ for at least one entry (i.e. the entry related to the reactant prefix of the corresponding transition). Thus the capacity to undertake some action, and therefore the corresponding events much differ in the two processes and their corresponding states. This ensures that the mapping $s \mapsto E(s)$ is a function. Moreover it is clear that $E(s_0) = e_0$ as required.

¹ cf. the use of a single exponential waiting-time in the SSA [24] and in the DSSAs [5,12] modeling the minimum of all the possible exponential waiting-times.

The function C is the clock assignment function; it associates a clock with each active event. The value of this clock is given by the distribution assignment function, F . Consider first the case of an exponentially-timed clock c_i . Clearly $F(0) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$ as desired, since this is the distribution function of an exponential distribution whose parameter is the sum of the rates of the actions enabled to start in the process configuration corresponding to the state s' . For the case $c_{\alpha,j}$, we can again see that the conditions of a distribution are satisfied since $F(0) = H(0 - \sigma(\alpha)) = H(-\sigma(\alpha)) = 0$, by definition of the Heaviside function. Moreover $\lim_{x \rightarrow \infty} H(x) = \lim_{x \rightarrow \infty} H(x - \sigma(\alpha)) = 1$ since $H(y) = 1, \forall y > 0$.

The next state function N defines a distribution over states for exponentially-timed events and a unique state for deterministically-timed events. In general, it is possible that two or more deterministically timed clocks will expire at the same time, but by probabilistic arguments we can assume that in a state the single exponentially timed event will not expire at the same moment of any of the deterministically timed events. Thus we can handle these two cases separately. The case of the single exponentially-timed event is straightforward. The completion of the exponentially-timed event corresponds to the first of the possible actions \mathcal{A}^+ firing; which action this is will be chosen according to the relative rates of all the actions which could have started from process P_i .

In general, a subset of deterministically-timed events may fire. In the SLTS we would handle such a case by an arbitrary interleaving of actions, each producing an update on the state vector, i.e. for a transition $\langle \mathcal{T}, \sigma, P_i \rangle \xrightarrow{(\alpha^-, r_{\alpha}, \sigma(\alpha))} \langle \mathcal{T}', \sigma, P_j \rangle$ there will be an update vector v_{α}^p which records the effect of completion of the action α when it started in process P_k . Applying such updates in succession cannot be distinguished from applying them simultaneously as happens in the next state function, that is

$$N(\langle P_i \rangle, E^*) = \left(\langle P_i \rangle + \sum_{e_{\alpha,k} \in E^*} v_{\alpha}^p \right) \rightarrow 1$$

when $E^* \subseteq E^-$. \square

5. Relation between Bio-PEPAD and Bio-PEPA

In this section we prove theorems stating the correspondence between the semantics of Bio-PEPA and Bio-PEPAD. More precisely, we start by introducing a notion of *interchangeability* between Bio-PEPA processes and Bio-PEPAD process configurations. Through a series of results on interchangeable processes we show that the SLTS of a Bio-PEPA process can be obtained from the SLTS of the corresponding interchangeable Bio-PEPAD process configuration. Moreover we relate probabilities in the SLTS of such a process configuration with those in the SLTS of the Bio-PEPA process.

We start by defining the inverse of function μ , denoted as $\mu^{-1} : \mathcal{C} \rightarrow \mathcal{P}$ and used to transform a Bio-PEPAD process configuration into a Bio-PEPA process

$$\begin{aligned} \mu^{-1}((\alpha, \kappa) \text{ op } S) &= (\alpha, \kappa) \text{ op } S & \mu^{-1}(P_1 \bowtie_{\mathcal{L}} P_2) &= \mu^{-1}(P_1) \bowtie_{\mathcal{L}} \mu^{-1}(P_2) \\ \mu^{-1}(S_1 + S_2) &= S_1 + S_2 & \mu^{-1}(S(l, L)) &= S(l). \end{aligned}$$

As reasonably expected, function μ is not bijective, namely even if in μ a unique Bio-PEPAD process configuration corresponds to each Bio-PEPA process, the opposite is generally false. Indeed it is the case that $\forall L \in \mathcal{L}_{\mathcal{D}}. \mu^{-1}(S(l, L)) = S(l)$, which means that we lose information about the structure of L , namely the actions started and not yet completed in $S(l, L)$.

We want to concentrate on those processes for which it is reasonable to define a valid notion of interchangeability.

Definition. A Bio-PEPA process $P \in \mathcal{P}$ and a Bio-PEPAD process configuration $P_C \in \mathcal{C}$ are said to be *interchangeable* if and only if

$$\mu(P) = P_C \wedge \mu^{-1}(P_C) = P.$$

Note that if P and P_C are interchangeable, then by definition $\mu(P) = P_C$ and, consequently, all the lists appearing in P_C must be empty. Practically, in P_C there must be no uncompleted actions running. Intuitively, this definition is constrained by the structure of Bio-PEPA processes which cannot have concurrently running uncompleted actions. Alternatively we could have defined μ^{-1} only on empty lists, i.e. as $\mu^{-1}(S(l, [])) = S(l)$, but in this case μ^{-1} would not have been a total function.

Now we prove the following theorem on the relation of interchangeability.

Theorem 5.1. Let $\mathcal{I} = \{(P, P_C) \mid P \in \mathcal{P}, P_C \in \mathcal{C}, \mu(P) = P_C, \mu^{-1}(P_C) = P\}$, then

$$\forall (P, P_C) \in \mathcal{I}. \forall P' \in \mathcal{P}. P \xrightarrow{(\alpha, w)}_c P' \implies \exists P'_C, P''_C \in \mathcal{C}. \quad P_C \xrightarrow{(\alpha^+, w)}_{st} P'_C \xrightarrow{(\alpha^-, w)}_{co} P''_C \wedge (P', P''_C) \in \mathcal{I}.$$

Proof. First we state two auxiliary equalities

$$\text{pick } \alpha [(l, \kappa, \alpha, \text{op})] = (l, \kappa, \alpha, \text{op}) \quad \text{del } \alpha [(l, \kappa, \alpha, \text{op})] = [] \quad (2)$$

which hold for any $(l, \kappa, \alpha, \text{op})$. We proceed by structural induction on P .

- Let us consider $P \equiv ((\alpha, \kappa) \downarrow S)(l)$ which is interchangeable with $P_C \equiv ((\alpha, \kappa) \downarrow S)(l, [\])$; from P there exists a unique possible derivation $P \xrightarrow{(\alpha, [S: \downarrow(l, \kappa)])}_c S(l - \kappa) \equiv P'$ if $k \leq l \leq N$. With the same condition from P_C we can derive $P_C \xrightarrow{(\alpha^+, [S: \downarrow(l, \kappa)])}_{st} S(l - \kappa, [(l, \kappa, \alpha, \downarrow)]) \equiv P'_C$; among all the possible derivations from P'_C there is one such that $P'_C \xrightarrow{(\alpha^-, [S: \downarrow(l, \kappa)])}_{co} S(l - \kappa, [\]) \equiv P''_C$ by Eq. (2). Finally, $(P', P''_C) \in \mathcal{I}$ since $\mu(S(l - \kappa)) = S(l - \kappa, [\]) \equiv P''_C$.
- Let us consider $P \equiv ((\alpha, \kappa) \uparrow S)(l)$ and $P_C \equiv ((\alpha, \kappa) \uparrow S)(l, [\])$; by similar arguments to the previous case, we have $P \xrightarrow{(\alpha^+, [S: \uparrow(l, \kappa)])}_c (\alpha, \kappa) \uparrow S(l + \kappa)$ if $0 \leq l \leq N - \kappa$; we have also $P_C \xrightarrow{(\alpha^+, [S: \uparrow(l, \kappa)])}_{st} S(l, [(l, \kappa, \alpha, \uparrow)]) \equiv P'_C$ and $P'_C \xrightarrow{(\alpha^-, [S: \uparrow(l, \kappa)])}_{co} S(l + \kappa, [\]) \equiv P''_C$ by Eq. (2). Finally, $(P', P''_C) \in \mathcal{I}$ since $\mu(S(l + \kappa)) = S(l + \kappa, [\]) \equiv P''_C$.
- Let us consider $P \equiv ((\alpha, \kappa) op S)(l)$ and $P_C \equiv ((\alpha, \kappa) op S)(l, [\])$ with $op = \oplus, \ominus, \odot$; by similar arguments to the previous cases we have $P \xrightarrow{(\alpha, [S: op(l, \kappa)])}_c (\alpha, \kappa) op S(l)$ with the appropriate condition on levels depending on op ; we have also $P_C \xrightarrow{(\alpha^+, [S: op(l, \kappa)])}_{st} S(l, [(l, \kappa, \alpha, op)]) \equiv P'_C$ and $P'_C \xrightarrow{(\alpha^-, [S: op(l, \kappa)])}_{co} S(l, [\]) \equiv P''_C$ by Eq. (2). Finally, $(P', P''_C) \in \mathcal{I}$ since $\mu(S(l)) = S(l, [\]) \equiv P''_C$.
- Let us consider $P \equiv (S_1 + S_2)(l)$ and $P_C \equiv (S_1 + S_2)(l, [\])$; we have that $P \xrightarrow{\alpha, w}_c S'_1(l') \equiv P'$ if $S_1(l) \xrightarrow{\alpha, w}_c S'_1(l')$. We assume the inductive hypothesis on $S_1(l)$ which means that $(S_1(l), P_{S_1}(l, [\])) \in \mathcal{I}$, $S_1(l) \xrightarrow{(\alpha, w)}_s S'_1(l')$ and $(S'_1(l'), P_{S'_1}(l', [\])) \in \mathcal{I}$. By considering the Bio-PEPAD semantics we have that $P_C \xrightarrow{(\alpha^+, w)}_{st} S'_1(l', L) \equiv P'_C \xrightarrow{(\alpha^-, w)}_{co} S'_1(l', [\])$ if $S_1(l, [\]) \xrightarrow{(\alpha^+, w)}_{st} S'_1(l', L)$ and $S'_1(l', L) \xrightarrow{(\alpha^-, w)}_{co} S'_1(l', [\])$. By applying the inductive hypothesis and noticing that L must contain only one element, we can apply Eq. (2). We then have that $P_{S_1}(l, [\]) \equiv S_1(l, [\])$, $P'_C = S'_1(l', L)$ and $P''_C \equiv S'_1(l', [\]) \equiv P_{S'_1}(l')$. The case in which S_2 makes the transition is symmetric.
- The case for $P \equiv C(l)$ comes from the inductive hypothesis on $S(l)$ once we apply $C \stackrel{def}{=} S$; namely $P' \equiv S'(l')$, $P_C \equiv S(l, [\])$ and $P''_C \equiv S'(l', [\])$.
- Let us consider $P \equiv P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha, w)}_c P'_1 \boxtimes_{\mathcal{L}} P_2$ if $\alpha \notin \mathcal{L}$ and $P_C \equiv \mu(P_1 \boxtimes_{\mathcal{L}} P_2)$. Note that $\mu(P_C) = \mu(P_1) \boxtimes_{\mathcal{L}} \mu(P_2)$ by definition of μ . We have to prove that there exist $P'_C, P''_C \in \mathcal{C}$ such that: $\mu(P_1 \boxtimes_{\mathcal{L}} P_2) \xrightarrow{(\alpha^+, w)}_{st} P'_C, P'_C \xrightarrow{(\alpha^-, w)}_{co} P''_C$ and $(P'_1 \boxtimes_{\mathcal{L}} P_2, P''_C) \in \mathcal{I}$. Note that to derive $P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha, w)}_c P'_1 \boxtimes_{\mathcal{L}} P_2$ we have that $P_1 \xrightarrow{(\alpha, w)}_c P'_1$. Now, since $(P_1, \mu(P_1)) \in \mathcal{I}$ then by inductive hypothesis on P_1 if $P_1 \xrightarrow{(\alpha, w)}_c P'_1$ then there exists $P'''_C \in \mathcal{C}$ such that $\mu(P_1) \xrightarrow{(\alpha^+, w)}_{st} P'''_C \xrightarrow{(\alpha^-, w)}_{co} \mu(P'_1)$ where $(P'_1, \mu(P'_1)) \in \mathcal{I}$. But since $\mu(P_1) \xrightarrow{(\alpha^+, w)}_{st} P'''_C$ we also have $\mu(P_1) \boxtimes_{\mathcal{L}} \mu(P_2) \xrightarrow{(\alpha^+, w)}_{st} P'''_C \boxtimes_{\mathcal{L}} \mu(P_2)$ when $\alpha \notin \mathcal{L}$. This then gives that $P'_C \equiv P'''_C \boxtimes_{\mathcal{L}} \mu(P_2)$. Moreover, since $P'''_C \xrightarrow{(\alpha^-, w)}_{co} \mu(P'_1)$ we also have $P'''_C \boxtimes_{\mathcal{L}} \mu(P_2) \xrightarrow{(\alpha^-, w)}_{co} \mu(P'_1) \boxtimes_{\mathcal{L}} \mu(P_2)$ when $\alpha \notin \mathcal{L}$. This then gives that $P''_C \equiv \mu(P'_1) \boxtimes_{\mathcal{L}} \mu(P_2) = \mu(P'_1 \boxtimes_{\mathcal{L}} P_2)$ and by definition of \mathcal{I} it holds that $(P'_1 \boxtimes_{\mathcal{L}} P_2, \mu(P'_1 \boxtimes_{\mathcal{L}} P_2)) \in \mathcal{I}$ and by definition of μ it holds that $(P'_1 \boxtimes_{\mathcal{L}} P_2, \mu(P'_1) \boxtimes_{\mathcal{L}} \mu(P_2)) \in \mathcal{I}$. The case in which P_2 makes the transition is symmetric. Moreover, the case in which both P_1 and P_2 move is a generalization of this case with two inductive hypotheses. \square

This theorem states a constructive result for the interchangeable relation stating that if P and P_C are interchangeable, then for any possible action derivable from P and leading to a state P' , there exists a sequence of start and completion transitions, from P_C through P'_C to P''_C , such that P' and P''_C are interchangeable. Thus we can think of interchangeability as a *simulation*, since everything which can be done from a Bio-PEPA process can be done by the interchangeable Bio-PEPAD process configuration. However note that P'_C is not interchangeable to P' since $\mu(P') \neq P'_C$ and it is fairly easy to see that $\nexists P'' \in \mathcal{P}$ such that P'' and P'_C are interchangeable, which follows by the fact that P'_C has a non-empty scheduling list and by the definition of interchangeability. This confirms our intuition that interchangeability is not a *bisimulation* as it is not symmetric. In particular,

$$\exists(P, P_C) \in \mathcal{I}. \exists P'_C, P''_C \in \mathcal{C}. P_C \xrightarrow{(\alpha^+, w)}_{st} P'_C \xrightarrow{(\alpha^-, w)}_{co} P''_C \wedge \nexists P'' \in \mathcal{P}. (P', P''_C) \in \mathcal{I}.$$

Let us denote a generic Bio-PEPA system $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P \rangle$ as $\langle \mathcal{T}, P \rangle$ whenever we are not concerned with the components of the system itself, and let us do the same for Bio-PEPAD system $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, \sigma, P \rangle$ by writing $\langle \mathcal{T}, \sigma, P \rangle$. We can now prove the following theorem.

Theorem 5.2. For any Bio-PEPA system $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, P \rangle$ there exists $P_C \in \mathcal{C}.(P, P_C) \in \mathcal{I}$ such that

$$\begin{aligned} \forall P' \in \mathcal{P}. \langle \mathcal{T}, P \rangle &\xrightarrow{(\alpha, r)}_s \langle \mathcal{T}, P' \rangle \implies \forall \sigma \in \Delta. \langle \mathcal{T}, \sigma, P_C \rangle \xrightarrow{(\alpha^+, r, \sigma(\alpha))}_s \langle \mathcal{T}, \sigma, P'_C \rangle \\ &\wedge \langle \mathcal{T}, \sigma, P'_C \rangle \xrightarrow{(\alpha^-, r, \sigma(\alpha))}_s \langle \mathcal{T}, \sigma, P''_C \rangle \\ &\wedge (P', P''_C) \in \mathcal{I}. \end{aligned}$$

Proof. We assume $\langle \mathcal{T}, P \rangle \xrightarrow{(\alpha, r)}_s \langle \mathcal{T}, P' \rangle$, which means that $P \xrightarrow{(\alpha, w)}_c P'$ and $r = f_\alpha[w, \mathcal{N}, \mathcal{K}]h^{-1}$. Using this we apply Theorem 5.1 and we have that $P_C \equiv \mu(P)$ and $P'_C \equiv \mu(P')$. We have that $\mu(P) \xrightarrow{(\alpha^+, w)}_{st} P'_C$ which means that we derive $\langle \mathcal{T}, \sigma, \mu(P) \rangle \xrightarrow{(\alpha^+, r_\alpha, \sigma(\alpha))}_s \langle \mathcal{T}, \sigma, P'_C \rangle$. Moreover, $P'_C \xrightarrow{(\alpha^-, w)}_{co} \mu(P')$ meaning that we also derive $\langle \mathcal{T}, \sigma, P'_C \rangle \xrightarrow{(\alpha^-, r_\alpha, \sigma(\alpha))}_s \langle \mathcal{T}, \sigma, \mu(P') \rangle$. Finally, we have that $(P', \mu(P')) \in \mathcal{I}$, which concludes the proof. \square

This theorem extends the notion of interchangeability to systems in a natural way. More precisely, if two processes are interchangeable, then any of the possible Bio-PEPA systems is interchangeable to an infinity of different Bio-PEPAD systems. This happens since any Bio-PEPAD system simulates the Bio-PEPA system, independently of the delays.

After these general results on interchangeability we can easily notice that there always exists a configuration interchangeable to any Bio-PEPA process, moreover such a configuration is unique. More formally, each process is interchangeable uniquely with the process configuration obtained by applying μ , namely $\forall P \in \mathcal{P}. (P, \mu(P)) \in \mathcal{I}$, and this is easily verifiable since by definition $\mu^{-1}(\mu(P)) = P$.

This means that we can build the Bio-PEPA stochastic semantics of a process P , namely its SLTS, by considering the semantics of a generic Bio-PEPAD system starting in $\mu(P)$ and traversing only interchangeable configurations. To clarify this intuition notice that all the states of the form $(n_1, n_2):0$ appearing in Fig. 7, namely $(3, 0):0$, $(2, 1):0$, $(1, 2):0$ and $(0, 3):0$ are interchangeable to the states $(3, 0)$, $(2, 1)$, $(1, 2)$ and $(0, 3)$ in Fig. 2, as stated by the previous theorem. We remark that this outlines a clear semantic relationship between the Bio-PEPAD system and the equivalent non-delayed Bio-PEPA system. We state the following corollary which formally characterizes the Bio-PEPA stochastic relation by means of Bio-PEPAD stochastic relation.

Corollary 5.3. The Bio-PEPA stochastic relation \rightarrow_s is equivalently defined by the following inference rule

$$\frac{\begin{array}{c} \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, \mu(P) \rangle \xrightarrow{(\alpha^+, r_\alpha, \sigma(\alpha))}_s \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, P'_C \rangle \\ \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, P'_C \rangle \xrightarrow{(\alpha^-, r_\alpha, \sigma(\alpha))}_s \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, \sigma, \mu(P') \rangle \end{array}}{\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, P \rangle \xrightarrow{(\alpha, r_\alpha)}_s \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, P' \rangle}$$

where σ is a generic function from Δ .

Proof. The proof comes from noting that Theorem 5.2 states a strong relationship between the stochastic derivations of Bio-PEPA processes and the corresponding stochastic derivations of Bio-PEPAD process configurations. More precisely, we rephrase Theorem 5.2 considering $(P, \mu(P)) \in \mathcal{I}$ so we have that, for any Bio-PEPA system $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, \text{Comp}, P \rangle$ there exists a Bio-PEPAD process configuration $\mu(P)$ such that P and $\mu(P)$ are interchangeable and any stochastic derivation from $\langle P \rangle$ to $\langle P' \rangle$ for action α and rate r , can be equivalently described by two stochastic derivations from $\langle \sigma, \mu(P) \rangle$ through a configuration $\langle \sigma, P'_C \rangle$ to $\langle \sigma, \mu(P') \rangle$ for the same action α and the same rate r . \square

We can apply these results and definitions to the toy examples discussed earlier in the paper. In particular, we have the interchangeability described by the following set

$$\begin{aligned} \mathcal{I} = \{ &(A(3) \boxtimes_{\{\alpha\}} B(0), A(3, [\])) \boxtimes_{\{\alpha\}} B(0, [\]), (A(2) \boxtimes_{\{\alpha\}} B(1), A(2, [\])) \boxtimes_{\{\alpha\}} B(1, [\])) \\ &(A(1) \boxtimes_{\{\alpha\}} B(2), A(1, [\])) \boxtimes_{\{\alpha\}} B(2, [\]), (A(0) \boxtimes_{\{\alpha\}} B(3), A(0, [\])) \boxtimes_{\{\alpha\}} B(3, [\])) \}. \end{aligned}$$

As a consequence, the SLTS of the Bio-PEPA process shown in Fig. 2 can be obtained by applying results of Theorem 5.3. Notice that this result could also be used to relate techniques for Bio-PEPA model checking to model checking of Bio-PEPAD process configurations. However, the embedding of the Bio-PEPA SLTS in the Bio-PEPAD one implies that model-checking Bio-PEPAD systems will generally be less tractable.

Theorem 5.2 relates Bio-PEPA semantics and Bio-PEPAD semantics. A final point relating to probabilities is worth discussing. We know that Bio-PEPAD models can be simulated by the DDA or by analysis techniques based on GSMPs. Thus we might investigate, for instance, what is the probability of observing in a DDA simulation a sequence of configuration changes such that the configurations are interchangeable to some processes. More precisely, given $P_C \equiv \mu(P)$ and $P'_C \equiv \mu(P')$ we aim to derive an analytical formula for the probability of the stochastic derivations

$$\forall \alpha \in \mathcal{A}. \langle \sigma, \mu(P) \rangle \xrightarrow{(\alpha^+, r_\alpha, \sigma(\alpha))}_s \langle \sigma, P'_C \rangle \xrightarrow{(\alpha^-, r_\alpha, \sigma(\alpha))}_s \langle \sigma, \mu(P') \rangle.$$

As we know, for each configuration there is a corresponding vector in the state space, so we have that $\langle \mu(P) \rangle = \mathbf{x}$, $\langle P_C \rangle = \mathbf{x} + \mathbf{v}_\alpha^r$ and $\langle \mu(P') \rangle = \mathbf{x} + \mathbf{v}_\alpha^r + \mathbf{v}_\alpha^p = \mathbf{x} + \mathbf{v}_\alpha$ where \mathbf{v}_α^r and \mathbf{v}_α^p denote the stoichiometry vector for the reactants and the products and are such that $\mathbf{v}_\alpha = \mathbf{v}_\alpha^r + \mathbf{v}_\alpha^p$. The probability $p(\mathbf{x})$ of observing equivalent state changes (i.e. \mathbf{x} modified into $\mathbf{x} + \mathbf{v}_\alpha^r$ modified into $\mathbf{x} + \mathbf{v}_\alpha$) is given in the DDA by the quantity

$$p(\mathbf{x}) = \sum_{i=1}^m \frac{a_i(\mathbf{x})}{a_0(\mathbf{x})} e^{-a_0(\mathbf{x} + \mathbf{v}_i^r)\sigma_i} \quad (3)$$

if the system contains reactions $\{R_i \mid i = 1, \dots, m\}$ and the delay of reaction R_i is σ_i . This equation is derived according to the following arguments. When the system is in state \mathbf{x} at time t , the next value for $\tau \sim \text{Exp}(a_0(\mathbf{x}))$ is sampled and reaction R_j is chosen to fire with probability $a_j(\mathbf{x})/a_0(\mathbf{x})$; notice that no reactions are already scheduled in the system since $P_C \equiv \mu(P)$. Assuming we chose reaction R_α , the state is changed from \mathbf{x} to $\mathbf{x} + \mathbf{v}_\alpha^r$ and time is increased to $t + \tau$. In the next step, a new value for $\tau' \sim \text{Exp}(a_0(\mathbf{x} + \mathbf{v}_\alpha^r))$ is sampled: if $\tau' > \sigma_\alpha$ then the state changes to $\mathbf{x} + \mathbf{v}_\alpha$ and time to $t + \sigma_\alpha$, otherwise a new reaction is scheduled. Our target event is $\tau' > \sigma_\alpha$ which has probability $\exp(-a_0(\mathbf{x} + \mathbf{v}_\alpha^r)\sigma_\alpha)$. Since events are independent, if we generalize among all possible reactions we get Eq. (3).

If we consider Eq. (3) in systems where $\forall R_i. f_{\alpha_i}[w, \mathcal{N}, \mathcal{K}] = a_i(\mathbf{x}) = r_{\alpha_i}$, by considering the Bio-PEPA definitions of α -derivative and exit rate for a process [14,15] rephrased for Bio-PEPAD, we can write a probability which is logically equivalent to $p(\mathbf{x})$ for $\mu(P)$ as

$$\mathbb{P}(\mu(P)) = \sum_{i=1}^m \frac{f_{\alpha_i}[w, \mathcal{N}, \mathcal{K}]}{\text{ExitRate}(\mu(P))} e^{-\text{ExitRate}(\mu(P))\sigma(\alpha_i)}. \quad (4)$$

This is an interesting result relating Bio-PEPAD and Bio-PEPA probabilities in the stochastic regime since

$$\lim_{\sigma \rightarrow \infty} \mathbb{P}(\mu(P)) = 0 \quad \lim_{\sigma \rightarrow 0} \mathbb{P}(\mu(P)) = \sum_{i=1}^m \frac{f_{\alpha_i}[w, \mathcal{N}, \mathcal{K}]}{\text{ExitRate}(\mu(P))}$$

where $\sigma \rightarrow k$ means $\forall i = 1, \dots, m. \sigma(\alpha_i) \rightarrow k$. In particular, in the limit $\sigma \rightarrow 0$ Eq. (4) reduces to the probability of leaving P , in its associated CTMC.

The probability of all the possible paths which satisfy the interchangeability property is given as the closure of $\mathbb{P}(\mu(P))$. This is the probability of observing, during a simulation of a Bio-PEPAD model, a series of steps which correspond to the interchangeable Bio-PEPA process. So, for instance, for the toy example in Fig. 7, the probability of observing the sequence of state changes

$$(3, 0) : 0 \rightarrow (2, 0) : 1 \rightarrow (2, 1) : 0 \rightarrow (1, 1) : 1 \rightarrow (1, 2) : 0 \rightarrow (0, 2) : 1 \rightarrow (0, 3) : 0$$

which is conceptually equivalent to the non-delayed sequence

$$(3, 0) : 0 \rightarrow (2, 1) : 0 \rightarrow (1, 2) : 0 \rightarrow (0, 3) : 0$$

of Fig. 2 is given by $\mathbb{P}(\tau > \sigma', \tau' > \sigma')$ which, since $\tau \sim \text{Exp}(2k)$ and $\tau' \sim \text{Exp}(k)$ are independent, evaluates as $e^{-3k\sigma'}$. Note that here the unique reaction is chosen with probability 1.

6. A model of the cell cycle with delays

In this section we encode in Bio-PEPAD a model of the cell cycle with delays as presented in [1,12]. Such a model is obtained by simplifying a DDE model of tumor growth including the immune system response and a phase-specific drug able to alter the natural course of action of the cell cycle of the tumor cells [42].

The model of the cell cycle with delays has been analyzed in [1] in order to discuss two possible interpretations of delays in the delay stochastic simulation algorithms, a delay-as-duration approach and a purely delayed approach. In this section, we simply show how to encode that model in Bio-PEPAD and for a detailed analysis of the model we refer to that paper.

The cell cycle is a series of sequential events leading to cell replication via cell division. It consists of four phases: G_1 , S , G_2 and M . The first three phases (G_1 , S , G_2) are called interphase (I). In these phases, the main event which happens is the replication of DNA. In the last phase (M), called mitosis, the cell segregates the duplicated sets of chromosomes between daughter cells and then divides to form two new cells in their interphase. The duration of the cell cycle depends on the type of cell (e.g. a normal human cell takes approximately 24 h to perform a cycle). Cell death via apoptosis may happen in any phase of the cell cycle. In Fig. 9 the model is graphically represented.

The Bio-PEPAD model considers two populations of cells: T_i , the population of tumor cells during cell cycle interphase, and T_M , the population of tumor cells during mitosis. We consider four possible actions, α , β , γ and δ , one for each of the events that we want to model. In particular, action α models the passage from the interphase to the mitotic phase, with rate a_1 , β models the mitosis, with rate a_4 , γ the death of a cell in the interphase, with rate d_2 , and δ the death of a cell in the mitotic phase, with rate d_3 . All the rates in the model refer to mass action kinetics.

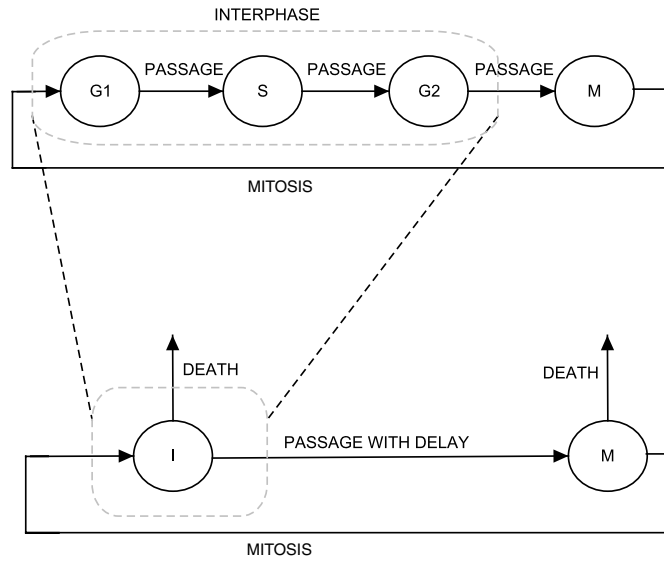


Fig. 9. The cell cycle model [1,12] and the role of the delay.

The Bio-PEPAd model is defined by the following species definitions:

$$T_I \stackrel{\text{def}}{=} (\alpha, 1)\downarrow + (\beta, 2)\uparrow + (\gamma, 1)\downarrow$$

$$T_M \stackrel{\text{def}}{=} (\alpha, 1)\uparrow + (\beta, 1)\downarrow + (\delta, 1)\downarrow$$

where the species behave as reactants or products, depending on their role as previously specified. Also, as all the actions obey a mass action kinetic law, we simply assume $f_\alpha = f_{MA}(a_1)$, $f_\beta = f_{MA}(a_4)$, $f_\gamma = f_{MA}(d_2)$ and $f_\delta = f_{MA}(d_3)$. The Bio-PEPAd process modeling the interactions is given by

$$T_I(n_0^I) \underset{\{\alpha, \beta\}}{\bowtie} T_M(n_0^M)$$

where n_0^I and n_0^M represent the initial concentration levels for the cells in the interphase and in the mitotic phase, respectively. Notice that γ and δ are not in the cooperation set since they model reactions involving a single species. Also, we note that this is also a valid Bio-PEPA process specification.

A delay $\sigma' > 0$ is used to model the duration of the interphase, hence the passage of a tumor cell from the population of those in the interphase to the population of those in the mitotic phase, namely the event modeled by action α , is delayed. To specify the delay in the Bio-PEPAd system to analyze, it is enough to define a function σ where

$$\sigma(\alpha) = \sigma' \quad \sigma(\beta) = \sigma(\gamma) = \sigma(\delta) = 0.$$

As a consequence, the Bio-PEPAd process initialized by applying function μ , namely the process configuration $T_I(n_0^I, [1]) \underset{\{\alpha, \beta\}}{\bowtie} T_M(n_0^M, [1])$, together with the function σ , completes the definition of the Bio-PEPAd system representing the cell cycle model.

By applying one of the techniques discussed in this paper this system can be analyzed. In particular, the Bio-PEPAd model can be automatically translated into a set of DDEs by applying the algorithm presented in Section 4.1. By computing the following vector of the kinetic laws

$$\begin{aligned} \overline{v_{KL}} &= (a_1 T_I(t - \sigma(\alpha)), a_4 T_M(t - \sigma(\beta)), d_2 T_I(t - \sigma(\gamma)), d_3 T_M(t - \sigma(\delta)))^T \\ &= (a_1 T_I(t - \sigma'), a_4 T_M(t), d_2 T_I(t), d_3 T_M(t))^T \end{aligned}$$

the following set of DDEs can be computed:

$$\frac{dT_I}{dt} = 2a_4 T_M - d_2 T_I - a_1 T_I(t - \sigma') \quad \frac{dT_M}{dt} = a_1 T_I(t - \sigma') - d_3 T_M - a_4 T_M.$$

As expected, this DDEs system is analogous to the one presented in [1]. The terms $d_2 T_I$ and $d_3 T_M$ represent cell deaths. The cells reside in the interphase at least σ' units of time; then the number of cells that enter mitosis at time t depends on the number of cells that entered the interphase σ' units of time before. This is modeled by the terms $T_I(t - \sigma')$ in the DDEs. Also, each cell leaving the mitotic phase produces two new cells in the T_I population, as given by terms $-a_4 T_M$ and $2a_4 T_M$. As a consequence, by defining the appropriate initial conditions for the resulting DDEs system it would be possible to reproduce the results presented in [1] for the deterministic model. In Fig. 10, taken from [1], the numerical solution of the DDEs in four

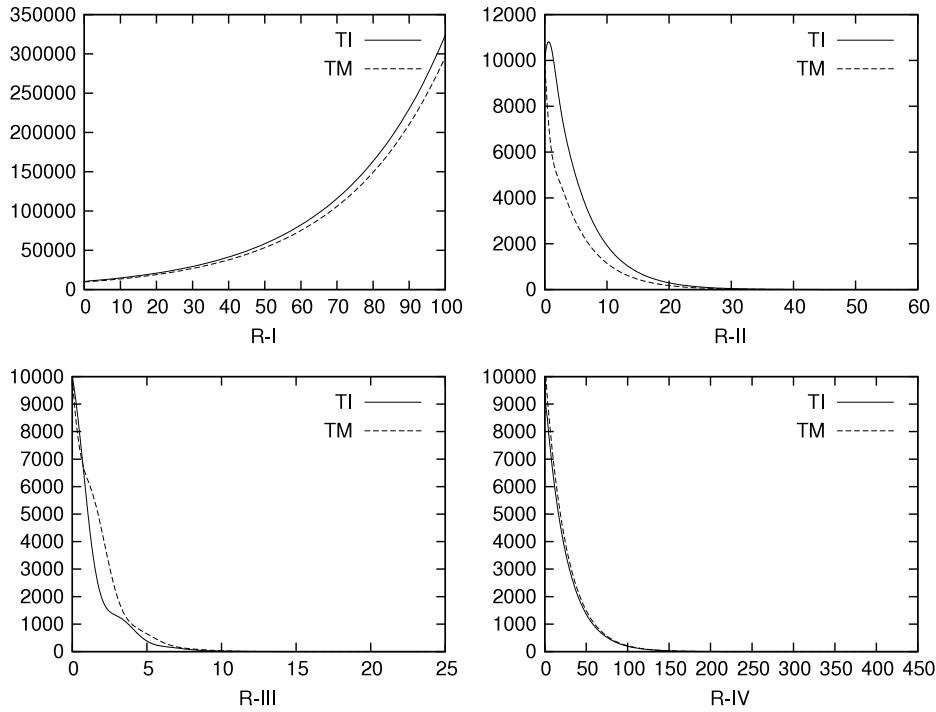


Fig. 10. Results of the numerical solution of the DDE model of the cell cycle with delay. On the x-axis time is given in days and on the y-axis is given the number of cells. The picture is taken from [1].

regions of parameters R-I, R-II, R-III and R-IV is shown. The parameters are as follows: $\sigma' = 1$, $a_4 = 0.5$ and $d_2 = 0.3$ in all the regions, $a_1 = 0.6$ and $d_3 = 0.1$ in R-I, $a_1 = 0.4$ and $d_3 = 0.5$ in R-II, $a_1 = 1$ and $d_3 = 1.3$ in R-III, $a_1 = 0.8$ and $d_3 = 0.3$ in R-IV. The initial state is $T_I(t) = T_M(t) = 10^5$ for $t \leq 0$.

As far as the stochastic analysis of the Bio-PEPAD systems is concerned, we can notice that the system we defined corresponds to the following set of reactions



where ϵ denotes the empty multi-set of products. Again, this is exactly the same reaction-based model used in [1] to compare the deterministic and the stochastic models for the cell cycle. Consequently, by applying the DSSA as explained in Section 4.2, it would be possible to reproduce the results presented in [1] for the stochastic model. In Fig. 11, taken from [1], the result of a single stochastic simulation of the system is shown for each one of the regions R-I, R-II, R-III and R-IV. The parameters are the same used to perform the analysis of the DDE model given in Fig. 10. In the figure the zoom on the time of eradication t_{er} (i.e. $T_I(t_{er}) = T_M(t_{er}) = 0$), if any, is also shown. In [1] is also discussed the effect of varying σ' on the dynamics of the system.

7. Discussion and conclusions

In this paper, we have enriched the stochastic process algebra Bio-PEPA with the possibility of assigning delays to actions, yielding the definition of a new non-Markovian process algebra: Bio-PEPAD. The use of delays in biological systems is suitable to model events for which the underlying dynamics cannot be precisely observed. Also, delays can be used to abstract portions of systems, leading to a reduced state space for models. From this point of view Bio-PEPA, which is based on the idea of levels to tackle the problem of state space explosion, was an appropriate candidate for defining our algebra.

The algebra is based on the syntax of Bio-PEPA. Hence the definition of Bio-PEPAD systems with delays can be easily obtained by adding, to a Bio-PEPA system of the target model, the delay specifications.

The semantics of the firing for the actions with delays is the delay-as-duration approach [5,1,12], as presented in the definition of DSSAs. In future work, we may enrich Bio-PEPAD with the other interpretation of delays presented in [1,12], in order to have the purely delayed approach and its combination with the one we currently consider.

The semantics of the algebra has been given in the Starting–Terminating style [25,10]. This permits us to observe the start and the completion of an action as two separate events, as required by delays. In future work, we will consider equivalence relations for Bio-PEPAD systems and processes, as done in [22] for the Bio-PEPA ones.

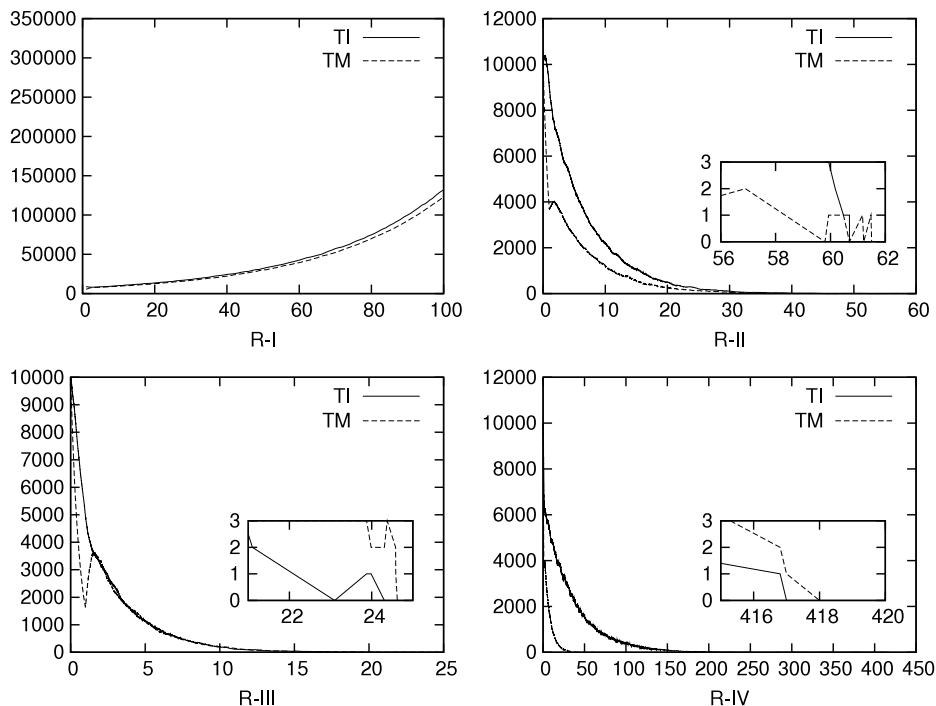


Fig. 11. Results of applying the DSSA with delay-as-duration approach to the model of the cell cycle with delay. On the x-axis time is given in days and on the y-axis is given the number of cells. The picture is taken from [1].

In keeping with the techniques developed for analyzing Bio-PEPA models, we showed the encoding of Bio-PEPAd systems in Generalized Semi-Markov Processes and we outlined how to perform stochastic simulation of Bio-PEPAd systems and how to automatically translate a Bio-PEPAd system in a set of Delay Differential Equations, the deterministic framework for the modeling of biological systems with delays. Moreover, the software framework for Bio-PEPA [7] can be extended to provide a tool for the automatic analysis of Bio-PEPAd systems.

In order to investigate the relation between Bio-PEPA and Bio-PEPAd systems, we proved results concerning the semantics of both the algebra. We introduced a notion of interchangeability of processes based on a notion of simulation, and showed how the semantics of Bio-PEPA systems can be given by the semantics of Bio-PEPAd ones. Moreover, we proved results on the probabilities of performing actions in the two algebras by investigating the probability of observing, during a simulation of a Bio-PEPAd system, a series of steps which correspond to the corresponding interchangeable Bio-PEPA system.

As far as applications of Bio-PEPAd are concerned, a well-known model of the cell-cycle where phase passages are abstracted by means of a delay has been discussed. We showed the translation of the Bio-PEPAd system modeling the cell-cycle into both a stochastic process with delays to be simulated by a DSSA, and a set of DDEs which can automatically derived by the system specification.

In the future, we plan to define Bio-PEPAd models of biological systems with delays and to analyze such models using the analysis techniques we defined in this paper. Moreover, equivalences for Bio-PEPAd systems will be defined and compared with existing equivalences for Bio-PEPA [22]. We think that, even in this case, the close relation between Bio-PEPA and Bio-PEPAd will allow us to naturally extend the theory that has already been developed. Finally, an interesting area for further future work will be to compare Bio-PEPAd with non-Markovian Stochastic Petri Nets such as DSPN [23].

Acknowledgements

The authors would like to thank the reviewers for their comments that helped to improve the manuscript.

References

- [1] R. Barbuti, G. Caravagna, A. Maggiolo-Schettini, P. Milazzo, On the interpretation of delays in delay stochastic simulation of biological systems, in: R. Back, I. Petre, E.P. de Vink (Eds.), *Proceedings of the 2nd Int. Workshop on Computational Models for Cell Processes, CompMod'09, EPTCS vol. 6* (2009), 17–29.
- [2] R. Barbuti, G. Caravagna, A. Maggiolo-Schettini, P. Milazzo, Delay stochastic simulation of biological systems: a purely delayed approach, *Trans. Comp. Sys. Bio.* XIII 6575 (2011) 61–84.
- [3] R. Barbuti, G. Caravagna, A. Maggiolo-Schettini, P. Milazzo, G. Pardini, The Calculus of Looping Sequences, in: M. Bernardo, P. Degano, G. Zavattaro (Eds.), *Formal Methods for Computational Systems Biology, SFM 2008, LNCS vol. 5016* (2008), 387–423.

- [4] R. Barbuti, A. Maggiolo-Schettini, P. Milazzo, G. Pardini, Spatial calculus of looping sequences, in: Int. Workshop From Biology to Concurrency and Back, FBTC'08, ENTCS 229 (1) (2008) 21–39.
- [5] M. Barrio, K. Burrage, A. Leier, T. Tian, Oscillatory regulation of Hes1: discrete stochastic delay modelling and simulation, *PLoS Comp. Bio.* 2 (9) (2006).
- [6] E. Beretta, T. Hara, W. Ma, Y. Takeuchi, Permanence of an SIR epidemic model with distributed time delays, *Tohoku Math. J.* 54 (2) (2002) 581–591.
- [7] Bio-PEPA web site <http://biopepa.org/>.
- [8] D. Bratsun, D. Volfson, L.S. Tsimring, J. Hasty, Delay-induced stochastic oscillations in gene regulation, *PNAS* 102 (41) (2005) 14593–14598.
- [9] M. Bravetti, Specification and analysis of stochastic real-time systems, Ph.D. Thesis, Università di Bologna, 2002.
- [10] M. Bravetti, R. Gorrieri, The theory of interactive generalized semi-Markov processes, *Theoret. Comput. Sci.* 282 (1) (2002) 5–32.
- [11] M. Bravetti, M. Bernardo, R. Gorrieri, Towards performance evaluation with general distributions in process algebras, in: Proc. of CONCUR'98, LNCS vol. 1466 (1998) 405–422.
- [12] G. Caravagna, Formal modeling and simulation of biological systems with delays, Ph.D. Thesis, Università di Pisa, 2011.
- [13] G. Caravagna, J. Hillston, Modeling biological systems with delays in Bio-PEPA, in: Proc. of the 4-th Workshop on Membrane Computing and Biologically Inspired Process Calculi, MeCBIC 2010, EPTCS vol. 40 (2010) 85–101.
- [14] F. Ciocchetta, J. Hillston, Bio-PEPA: a framework for the modelling and analysis of biochemical networks, *Theoret. Comput. Sci.* 410 (33–34) (2009) 3065–3084.
- [15] F. Ciocchetta, J. Hillston, Calculi for biological systems, in: M. Bernardo, P. Degano, G. Zavattaro (Eds.), *Formal Methods for Computational Systems Biology*, SFM 2008, LNCS vol. 5016 (2008), 265–312.
- [16] D.R. Cox, The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables, *Proc. Cambridge Phil. Soc.* 51 (1955) 433–440.
- [17] R.V. Culshaw, S. Ruan, A Delay-differential equation model of HIV infection of CD4+ T-cells, *Math. Biosci.* 165 (2000) 27–39.
- [18] P.R. D'Argenio, J.-P. Katoen, A theory of stochastic systems, part II: process algebra, *Inf. and Comp.* 203 (1) (2005) 39–74.
- [19] P.R. D'Argenio, J.-P. Katoen, E. Brinksma, A stochastic automata model and its algebraic approach, in: Proc. 5th Int. Workshop on Process Algebra and Performance Modeling, PAFM'97, CTIT technical reports series 97–14, University of Twente, 1997, pp. 1–16.
- [20] V. Danos, J. Feret, W. Fontana, R. Harmer, J. Krivine, Rule-based modelling of cellular signalling, in: Proc. of CONCUR'07, LNCS vol. 4703 (2007) 17–41.
- [21] L. Dematté, C. Priami, A. Romanel, Modelling and simulation of biological processes in BlenX, *SIGMETRICS Perform. Eval. Rev.* 35 (4) (2008) 32–39.
- [22] V. Galpin, J. Hillston, A semantic equivalence for Bio-PEPA based on discretisation of continuous values, *Theoret. Comput. Sci.* 412 (21) (2011) 2142–2161.
- [23] R. German, *Performance Analysis of Communication Systems with Non-Markovian Stochastic Petri Nets*, John Wiley & Sons, Inc., 2000.
- [24] D. Gillespie, Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem.* 81 (1977) 2340.
- [25] R.J. van Glabbeek, F.W. Vaandrager, Petri net models for algebraic theories of concurrency, *LNCS* 259 (1987) 224–242.
- [26] P.W. Glynn, On the role of generalized semi-markov processes in simulation output analysis, in: Proc. of the 15th Conference on Winter Simulation, Volume 1, 1983, pp. 39–44.
- [27] J. Hillston, *A Compositional Approach to Performance Modelling*, Cambridge University Press, 1996.
- [28] J. Heath, M. Kwiatkowska, G. Norman, D. Parker, O. Tymchyshyn, Probabilistic model checking of complex biological pathways, *Theoret. Comput. Sci.* 391 (2008) 239–257.
- [29] N. Lopez, M. Nunez, NMSPA: a non-Markovian model for stochastic processes, in: T.-H. Lai (Ed.), Proc. of ICDS 2000, IEEE, 2000, pp. 33–40.
- [30] J. Markovski, Real and stochastic time in process algebras for performance evaluation, Ph.D. Thesis, University of Eindhoven, 2008.
- [31] P. Milazzo, Qualitative and quantitative formal modeling of biological systems, Ph.D. Thesis, Università di Pisa, 2007.
- [32] R. Milner, M. Toft, R. Harper, *The Definition of Standard ML*, MIT Press, 1990.
- [33] I. Mura, D. Prandi, C. Priami, A. Romanel, Exploiting non-Markovian bio-processes, *ENTCS* 253 (3) (2009) 83–98.
- [34] L. Paulevė, S. Youssef, M.R. Lakin, A. Phillips, A generic abstract machine for stochastic process calculi, in: *Proceedings of the 8th International Conference on Computational Methods in Systems Biology CMSB* vol. 10, 2010, pp. 43–54.
- [35] M. Pedersen, G. Plotkin, A language for biochemical systems: design and formal specification, *Trans. Comp. Sys. Bio.* XII 5945 (3) (2010) 77–145.
- [36] G. Plotkin, The origins of structural operational semantics, *J. Log. Alg. Progr.* 60–61 (2004) 3–15.
- [37] C. Priami, Stochastic π -calculus, *The Comp. J.* 38 (7) (1995) 578–589.
- [38] C. Priami, A. Regev, E. Shapiro, W. Silverman, Application of a stochastic name-passing calculus to representation and simulation of molecular processes, *Inform. Process. Lett.* 80 (2001) 25–31.
- [39] A. Regev, E.M. Panina, W. Silverman, L. Cardelli, E.Y. Shapiro, Bioambients: an abstraction for biological compartments, *Theoret. Comput. Sci.* 325 (1) (2004) 141–167.
- [40] A. Regev, W. Silverman, E. Shapiro, Representation and simulation of biochemical processes using the pi-calculus process algebra, in: *Pacific Symposium on Biocomputing*, vol. 6, World Scientific Press, 2001, pp. 459–470.
- [41] Prism web site <http://www.prismmodelchecker.org/>.
- [42] M. Villasana, A. Radunskaya, A delay differential equation model for tumor growth, *J. Math. Biol.* 47 (2003) 270–294.
- [43] F. Zhang, Z. Lia, F. Zhang, Global stability of an SIR epidemic model with constant infectious period, *Appl. Math. Comp.* 199 (1) (2008) 285–291.